UNIVERSITY OF SZEGED
DOCTORAL SCHOOL OF EDUCATION

IJTIHADI KAMILIA AMALINA

ASSESSMENT OF STUDENTS' MATHEMATICAL PROBLEM-SOLVING SKILLS AND THE FACTORS INFLUENCING THEM

DOCTORAL DISSERTATION

SUPERVISOR: PROFESSOR DR. TIBOR VIDÁKOVICH

SZEGED, HUNGARY, 2025

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my supervisor, Professor Tibor Vidákovich, for his continuous support, motivation, guidance, patience, and vast knowledge during my Ph.D. study, research, and publications.

I extend my gratitude to the Hungarian government and the Tempus Public Foundation for granting me the Stipendium Hungaricum scholarship for my Ph.D. study. The scholarship made my dream of studying at the University of Szeged come true.

My sincere thanks to my family, my mother and my sisters for the endless support both spiritually and mentally, encouragement, and unconditional sacrifice during my Ph.D.

I would like to thank the professors, lecturers, and staff of the Doctoral School of Education, University of Szeged, for their valuable courses and unwavering support. I also would like to thank my classmates and the Doctoral School of Education students for the impressive discussions and suggestions. I thank my fellows in Szeged for the support and fun we have had in the past years.

I would like to thank my professors, colleagues, and alumni of Universitas Negeri Surabaya for their help during my data collections. I would like to thank the principals, teachers, and students for their help and participation in my dissertation project.

I extend my thanks to the University of Szeged and the Klebelsberg Library for useful facilities and academic support. Last but not least, I would like to thank all the people behind my success that I could not list one by one.

DEDICATION

I dedicate my dissertation to my mom, Mudayanah– a supermom whose love and dedication have always been unwavering, even in the face of life’s challenges. Mom, as you always wish in your night prayer me to be a professional educator and researcher, and this dissertation is the beginning of that journey. I sincerely thank you for your endless love and encouragement.

TABLE OF CONTENTS

Contents

CHAPTER 1. GENERAL INTRODUCTION

1.1 Background and Rationale

Mathematics is a foundation to understand phenomena, the environment, and other disciplines to participate systematically in daily life (Anjum, 2015; Imam & Singh, 2015; Junpeng et al., 2020; Khairani & Sahari Nordin, 2011). For this reason, it is a compulsory subject at the primary and secondary education levels (Du Toit & Du Toit, 2013; Jones et al., 2015). Despite its importance, applying mathematics to real-world and professional contexts remains a challenge for graduate students because the mathematics curriculum is separated from practice-oriented mathematics (Jones et al., 2015; Ke & M. Clark, 2020; Shute et al., 2016). Hence, problem-solving has been incorporated into mathematics curricula as a core standard to bring contextual problems into the mathematics classroom (Du Toit & Du Toit, 2013; Jones et al., 2015; Karatas & Baki, 2013). However, students continue to face difficulties and have low performance in mathematical problem-solving (Greiff et al., 2013; OECD, 2019; Ramos et al., 2021; Scherer & Beckmann, 2014).

Some scholars address several ways to improve students' mathematical problem-solving skills. First, investigate profiles and development across grades to optimize the teaching method in their best period of development (Greiff et al., 2013; Molnár et al., 2013). Second, evaluate skill differences according to background variables to reach equality in mathematics education (Anjum, 2015; Williams, 2005). Lastly, examine the factors influencing mathematical problem-solving to design appropriate instructional strategies that consider these factors (Bahar, 2013; Csapó & Molnár, 2017; Scherer & Beckmann, 2014).

Despite these efforts, several drawbacks have been identified. These include limited information on the development of the skills in each problem-solving phase, inconsistent results of the outperformed subsample, a lack of studies in collectivist cultures, and frequent use of monodisciplinary tasks (Greiff et al., 2013; Grissom, 2004; Halpern et al., 2005; J. Lee & McIntire, 2000; Molnár et al., 2013; Nepal, 2017; Ramos et al., 2021; Williams, 2005). Additionally, some scholars have shown that mathematics domain-specific prior knowledge (DSPK), science knowledge, text comprehension, parents' education, and family income have important roles in shaping mathematical problem-solving skills (Boonen et al., 2014; Csapó & Molnár, 2017; Daroczy et al., 2020; Scherer & Beckmann, 2014; Vista, 2013). However, some limitations exist, such as analyzing these factors as unconnected factors, the lack of a relevant mathematics DSPK test, and the lack of studies on collectivist cultures.

One of the highlighted issues is that previous studies used a monodisciplinary task (Du Toit & Du Toit, 2013; İncebacak & Ersoy, 2016; Widodo et al., 2021). This task fails to address the challenge of the 21st century in a realistic interdisciplinary context by combining Science, Technology, Engineering, and Mathematics (STEM) (Jolly, 2016; Kelley & Knowles, 2016; Maass et al., 2019). Although interdisciplinary STEM problem-solving assessments have been developed, existing instruments tend to emphasize science education while underrepresenting mathematics (Annaggar & Tiemann, 2020; Lasa et al., 2020; Maass et al., 2019; Scherer & Tiemann, 2012; Shute et al., 2016). Additionally, issues exist regarding cultural bias, the practicality of classroom implementation, and the psychometric evidence of existing assessments (Amalina & Vidákovich, 2022b).

Another highlighted issue is the lack of studies on collectivist culture, particularly in

Indonesia (Ramos et al., 2021; Williams, 2005). Culture, specifically collectivist versus western orientations, impacts achievement in mathematics (Meng & Liu, 2022). In western cultures, education tends to emphasize active learning, student-centered approaches, intrinsic motivation, and individualism (Cortina et al., 2017; Niu et al., 2025). In contrast, collectivist cultures place greater emphasis on teacher-centered educational system, social connectedness, interdependence, and group-oriented goals, often driven by extrinsic motivation (Cortina et al., 2017; Leung, 2001; Niu et al., 2025). Culture also influences the contexts used in students' tasks, which in turn affects success in problem-solving (Arieli & Sagiv, 2018; Salgado, 2017).

Indonesia, as one of the largest collectivist countries, still follows the trend of monodisciplinary mathematical problem-solving assessment by focusing on a single grade, which fails to address 21st century challenges (Amalina & Vidákovich, 2022b). This trend is in contrast to the 2013 Indonesian mathematics curriculum, which requires middle school students to solve real-life problems through STEM integration and engineering design with technology (Government Regulation Number 24 Year of 2016 Attachment 15, 2016). Consequently, Indonesian students struggle with curriculum-based assessments due to the lack of appropriate assessment tools for daily instruction and their limited familiarity with such tasks (Suratno et al., 2020). The difficulties of Indonesian middle school students in solving mathematical problems are also shown in international assessments, such as the Program for International Student Assessment (PISA) 2018, where Indonesia ranked 72nd out of 78 countries (OECD, 2019).

Therefore, the main aim of the present study is to assess Indonesian middle school students' mathematical problem-solving skills using a STEM-based context task. It also aims to examine the cognitive and socioeconomic status (SES) factors influencing their skills using structural equation modeling (SEM). Prior to this assessment, a review study of the trend assessment in this area and the development of relevant tests is required. The test is developed with an emphasis on mathematics and the integration of STEM context, focusing on a classroom-friendly test that highlights process-based assessment.

By conducting this study, teachers and researchers, particularly in Indonesia, can design instructional programs targeting underrepresented groups. These programs can be implemented during the best period of the development of skills and focus on key influencing factors. As a result, this helps reduce students' difficulties and improve their performance in this area.

1.2 Literature Review

1.2.1 Mathematical Problem-Solving Skills in the STEM Context: Theoretical Foundations

Problem-solving is a cognitive or mental activity focused on transforming a given situation into a goal situation when there are no direct ways to reach the goal (Álvarez et al., 2019; Carifio, 2015; Di Martino & Signorini, 2019; Lester, 2013; Peter Liljedahl et al., 2016; Schoenfeld, 2013; Wilson et al., 1989). According to the definition, there are four main problem-solving keywords: cognitive activity (process), problematic situation (containing content), goal, and unknown solution method.

Mathematical problem-solving is a thinking process of using mathematical knowledge to understand the problem situation and to reach the goal (Nunokawa, 2005). Mathematical problem-solving has two important characteristics: the task and the process (Chamberlin, 2008;

Jäder et al., 2020). The process of mathematical problem-solving begins with exploring and understanding a problem, mathematizing a situation, creating assumptions and considering these assumptions in relation to the final solution, revising current knowledge, creating a new technique, defining a situation, applying mathematical knowledge, and evaluating the solution (Chamberlin, 2008; Jäder et al., 2020; Nunokawa, 2005).

The task in mathematical problem-solving is non-routine, authentic, uses realistic context, uses more than one tool and approaches, implements multiple algorithms, and promotes flexibility in thinking (Chamberlin, 2008; Jäder et al., 2020). The definition of the context in the problem-solving task refers to the description of the scenario in which the task is situated (Heuvel-Panhuizen, 2005; Salgado, 2017). The use of context is with the purpose of transferring mathematical knowledge or content into practice or the real world (Salgado, 2017). The relevant context of mathematical problem-solving is a realistic mathematics problem (Heuvel-Panhuizen, 2005). It presents a realistic scenario for students to understand and apply relevant mathematics concepts (Heuvel-Panhuizen, 2005; Salgado, 2017).

Integrating the realistic STEM context into mathematical problem-solving tasks addresses the needs and challenges of the 21st century (Jolly, 2016; Kelley & Knowles, 2016; Maass et al., 2019). Mathematical problem-solving in the context of integrated STEM is developed under the theory of STEM education (Jolly, 2016; Kelley & Knowles, 2016; Maass et al., 2019). STEM education serves as a bridge between realistic mathematics problems and problem-solving (Jurdak, 2016; Zeeshan et al., 2021). In other words, problem-solving acts as an integrative object between mathematics and STEM. One of the definitions of STEM education refers to applying the content and concepts of mathematics and science through engineering design methodology and using appropriate technology (Shaughnessy, 2013; Tuong et al., 2023; Zeeshan et al., 2021). For example, technology (T) and engineering design (E) are combined with inquiry (S) and applications of mathematics (M) in problem-solving (Jurdak, 2016). According to this definition, it is possible to use the STEM context to support the identity of mathematics education as a separate subject through problem-solving.

A mathematical problem in the context of integrated STEM means that mathematics serves as the core for solving the problem, while STEM provides the methodologies and context (Jurdak, 2016; Miteva et al., 2022; Vukelic et al., 2020). In this case, mathematics plays a role in the content used, the way of thinking, problem handling, modeling, representation, analysis, prediction, and reasoning (Jurdak, 2016; Lasa et al., 2020; Miteva et al., 2022; Shaughnessy, 2013; Vukelic et al., 2020). Science participates in the context used, methodology, and way of thinking through scientific inquiry (Miteva et al., 2022; Shaughnessy, 2013). Engineering and technology play a role in the methodology for connecting mathematics concepts and science through engineering-based design (Miteva et al., 2022; Shaughnessy, 2013; Vukelic et al., 2020; Zeeshan et al., 2021). In addition, technology also provides tools to ease the process of solving problems, both electronically and non-electronically (Jolly, 2016; Shaughnessy, 2013; Zeeshan et al., 2021).

Therefore, the definition of a mathematical problem in the integrated STEM context refers to a realistic mathematical problem for applying mathematical knowledge through the STEM context. Mathematics is at the core of the problem, serving both as content and as a way of thinking. Science plays a role in providing context and fostering inquiry-based thinking. Engineering participates as a methodology to solve the problem through engineering-based design. Technology serves as a tool to solve the problem. Hereinafter, the term integrated STEM-based mathematical problem-solving is used to define mathematical problem-solving in the integrated STEM context. The description of the definition of an integrated STEM-based mathematical problem in mathematical problem-solving is illustrated in Figure 1.1.

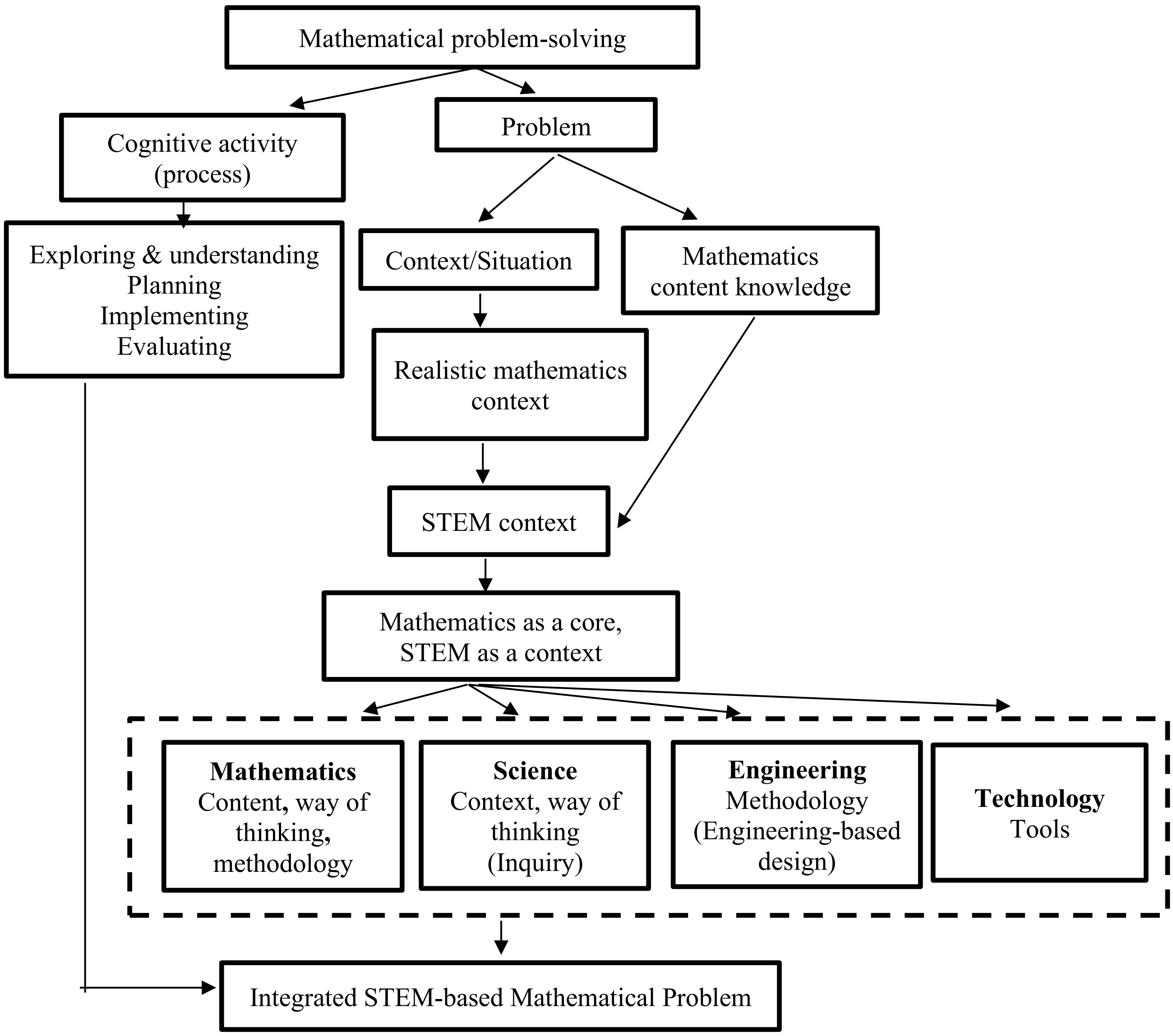


Figure 1. 1 The Definition of an Integrated STEM-Based Mathematical Problem in Mathematical Problem-Solving

1.2.2 STEM Integration and Mathematical Problem-Solving Skills in Indonesian Educational Policy

The Indonesian government has participated in addressing the need to integrate the STEM context into classrooms through the 2013 national curriculum, with the latest version in 2016 (Government Regulation Number 24 Year of 2016 Attachment 15, 2016). It emphasizes integrating the STEM context through problem-solving in mathematics and science subjects, particularly for middle school students. This is because middle school students begin to think critically with the fastest development skills.

All schools, including public and private schools with varying accreditation levels from low (C-accreditation) to high (A-accreditation), must implement the national curriculum. However, international public and private schools are able to combine the national curriculum with other curricula (e.g., Cambridge curriculum, etc.). The implementation of the curriculum is also one of the criteria used by the Indonesian Ministry of Education (MOE) to accredit schools, along with factors such as their systems, environment, human resources, and teaching and learning processes.

The national curriculum also has to be implemented at all levels of education. The educational level in Indonesia is divided into kindergarten (International Standard Classification of Education (ISCED) 0), primary school (ISCED 1), middle school (ISCED 2), and high school (ISCED 3). The kindergarten level is a 2-year program for children aged 4 to 6 years that emphasizes the development of spiritual, emotional, and foundational cognitive skills. The primary school level is 6-year education for students aged 6 to 12. The middle school level is a 3-year program after primary school. It leads to a 3-year program of either vocational or general high school program (ISCED 3). The general high school focuses on theoretical education and offers three main areas of study: science, social studies, and language. The vocational high school emphasizes practical education (Attachment of Government Regulation Number 21 Year of 2016, 2016).

Mathematics is a mandatory subject at primary and secondary education levels. Core mathematics is required for all students, while advanced mathematics is only required for those who choose to study science in high school. Problem-solving skills by integrating STEM contexts are obligatory mathematics skills at every level (Government Regulation Number 24 Year of 2016 Attachment 15, 2016; Government Regulation Number 24 Year of 2016 Attachment 6, 2016). It is determined through the core and basic competencies for the skill dimension. The skill dimension aims to apply knowledge that has been learned by students. Since problem-solving is mentioned as a compulsory skill, teachers have to infuse it into teaching and learning processes and assessments. Problem-solving as a core skill in the curriculum is supported by the teacher textbooks provided by the Indonesian MOE. These textbooks guide teachers with available activities containing real-world problems and their alternative solutions to teach concepts. Additionally, the Indonesian MOE provides assessments to assess problem-solving skills in the textbooks, but the scientific method to solve the problem is less emphasized.

Science, in the form of natural science, is only mandatory for students in grades 4 to 9. In high school, science courses are divided into physics, chemistry, and biology, and students choose their courses based on their area of interest. Technology is used as a tool in the teaching and learning process, and students need to apply engineering thinking in the process of solving a problem using a design-based process (Attachment of Government Regulation Number 21 Year of 2016, 2016; Government Regulation Number 24 Year of 2016 Attachment 15, 2016; Government Regulation Number 24 Year of 2016 Attachment 6, 2016). However, there is no engineering subject offered from kindergarten through middle school.

There are no differences between core competencies for sciences and mathematics in characters (spiritual & social-emotional), cognitive, and skills dimensions. The basic competencies of mathematics and science in each grade are different, with the highest grade having the most complicated abilities that have to be learned. However, there are anchor topics, competencies, and skills at every level (e.g., grades 7 to 9 at the middle school level).

1.2.3 Factors Influencing Students' Mathematical Problem-Solving Skills

Mathematical problem-solving requires complex skills and knowledge (Álvarez et al., 2019; Di Martino & Signorini, 2019). Therefore, several researchers have identified internal factors (from the students or individual) influencing mathematical problem-solving, including resources, heuristics and methods, control, and affect (belief, emotion, attitude, etc.) (Carlson & Bloom, 2005). Resources refer to formal and informal conceptual understanding, knowledge, facts, and procedures. Resources are also related to cognitive factors, including DSPK or prior knowledge, text or reading comprehension, creativity, memory, intelligence, verbal ability, spatial ability, and quantitative ability (Bahar, 2013; Männamaa et al., 2012; Swanson & Beebe-Frankenberger, 2004). Despite internal factors, some external factors have been identified as impacting mathematical problem-solving skills, such as demographic variables (e.g., gender, grade, school location, etc.), SES (e.g., family income, parents' occupations, parents' education, etc.), and characteristics of the task (Greiff et al., 2013; Molnár et al., 2013; Vilenius-Tuohimaa et al., 2008; Zhu, 2007).

Pangeni (2014) investigated factors influencing mathematics achievement and problem-solving, and found that the most influential factors are family (SES: parents' education, family income, etc.) and student (cognitive knowledge) characteristics, which accounted for 29.9% and 18% of variances, respectively. Therefore, it is essential to emphasize SES and cognition as potential factors impacting mathematical problem-solving skills.

Cognitive factors influencing mathematical problem-solving vary based on the structure and content of the problem (Bahar, 2013). Mathematical problem-solving requires understanding the problem situation or scenario using prior mathematical knowledge and applying relevant mathematical concepts (Nunokawa, 2005). Prior mathematical knowledge is also known as mathematics DSPK (Hailikari et al., 2007). It refers to conceptual knowledge (what and why a procedure is appropriate for a task) and procedural knowledge (how and when to use the appropriate procedures) (Al-Mutawah et al., 2019; Booth & Davenport, 2013; Hailikari et al., 2007). This foundational knowledge serves as a prerequisite for engaging in more complex skills, such as mathematical problem-solving skills (Hailikari, 2009).

Several scholars have highlighted that mathematics DSPK, both procedural and conceptual knowledge, has been proven to be an important factor in the success of mathematical problem-solving (Al-Mutawah et al., 2019; Bahar, 2013; Booth & Davenport, 2013; Byrnes & Takahira, 1994; Jitendra et al., 2013; C. Y. Lee & Chen, 2009; Serafino & Cicchelli, 2003).

Accordingly, the tests used to assess mathematics DSPK and mathematical problem-solving differ, as mathematics DSPK serves as a prerequisite for solving problems. However, the term "mathematical problem-solving test" has been used in some studies to describe assessments that actually focus on mathematics DSPK (Al-Mutawah et al., 2019; Süß & Kretzschmar, 2018). This is because these assessments include indicators related to problem-solving steps, such as applying basic procedures, even though they were designed to evaluate prior mathematical knowledge.

Despite mathematical knowledge, students must comprehend the language and text to understand the problem situation (Fuentes, 1998; Spencer et al., 2020). Therefore, the cognitive process of understanding the meaning of the text, also called text comprehension skills, is crucial for mathematical problem-solving (Timm & Uibu, 2015). Some scholars have found that text comprehension skills are among the most important factors influencing mathematical problem-solving (Boonen et al., 2014; Daroczy et al., 2020; Lin, 2021; Männamaa et al., 2012; Vilenius-Tuohimaa et al., 2008).

Integrated STEM-based mathematical problem-solving uses science as the context, methodology, and way of thinking (Miteva et al., 2022; Shaughnessy, 2013). Therefore, science knowledge also contributes to the success of the problem-solving process (Csapó & Molnár, 2017; Scherer & Beckmann, 2014). However, students' performance in science knowledge also depends on text comprehension skills (Akbasli et al., 2016; Cano et al., 2014). Hence, text comprehension skills also indirectly affect mathematical problem-solving through science knowledge.

Parents' education is related to parents' beliefs, parenting styles, and the support given to students (Alomar, 2007; Davis-Kean et al., 2021; Engin-Demir, 2009). Additionally, parents' education influences their family income (Alomar, 2007; Davis-Kean et al., 2021). Parents with higher income tend to provide more material resources (Davis-Kean et al., 2021). Therefore, parents' education and family income are found to be crucial factors influencing mathematical problem-solving skills (Alghazo & Alghazo, 2015; Davis-Kean et al., 2021; Kodippili, 2011; Pangeni, 2014).

1.3 Framework of Integrated STEM-Based Mathematical Problem-Solving

STEM is infused in the methodology and context of the problem. Therefore, integrating and combining several frameworks from each STEM discipline is crucial, as demonstrated in previous studies (Kelley & Knowles, 2016; Priemer et al., 2020; Wells, 2016). A well-formulated and comprehensive general problem-solving framework, such as the PISA creative problem-solving, was used as the ground theory. It includes exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting. Indicators for each phase in the PISA creative problem-solving were developed by combining problem-solving frameworks from each STEM discipline and eliminating duplicate indicators. Since the frameworks for general problem-solving, mathematical problem-solving, science problem-solving, and engineering problem-solving overlap, it is possible to combine these frameworks and eliminate duplicate indicators (Csapó & Molnár, 2017; Scherer & Beckmann, 2014).

Several fundamental mathematical problem-solving frameworks were used to develop an integrated STEM-based mathematical problem-solving framework. First, Polya's (1945) heuristic problem-solving model includes understanding a problem, devising a plan, carrying out the plan, and looking back. Second, Schoenfeld's (1985) model includes reading, analysis, exploration, planning, implementation, and verification. Lastly, the frameworks by Bayazit (2013) and Wu & Adams (2006) emphasize the application of mathematical concepts, procedures, rules, facts, mathematization, and standard computational skills during the carrying out phase. The term "hypothesis" in scientific inquiry is similar to the term "plan" in mathematical heuristic problem-solving (Priemer et al., 2020). Hence, making a hypothesis is also considered a part of mathematical problem-solving. Other mathematical problem-solving frameworks used include mathematics experimentation and inquiry-based mathematics education, which emphasize generating and testing hypotheses. A proving framework is also employed, including making a hypothesis, selecting a reasonable argument, and deducting, proofing, or finding a counter example (Boero, 1999; Priemer et al., 2020).

Scientific inquiry served as a science problem-solving framework used to develop indicators for an integrated STEM-based mathematical problem-solving framework. Indicators in scientific inquiry include investigating problems, observing, experimenting, measuring, testing, generating hypotheses, developing explanations and solutions, and evaluating (Priemer et al., 2020).

The engineering problem-solving framework incorporated into the development of an integrated STEM-based mathematical problem-solving framework was engineering-based design (Jolly, 2016; Kelley & Knowles, 2016; Wells, 2016). It aims to connect science and mathematical concepts. Baele (2017) identified engineering-based design steps as: (1) asking; (2) imagining; (3) planning; (4) creating; (5) testing; and (6) evaluating. Another researcher proposed a linear process of engineering-based design that includes communication (Robinson, 2016). Other engineering problem-solving frameworks, similar to heuristic problem-solving, suggest a linear model (Gray et al., 2005) and a cyclic model (Woods, 2000). These models are similar to scientific inquiry in certain phases (e.g., constituting a question, developing an empirical test, and making a conclusion).

According to the description of several mathematical, science, and engineering problem-solving frameworks applicable for developing an integrated STEM-based mathematical problem-solving framework, specific steps can be identified. The mathematical problem-solving steps in the integrated STEM-based mathematical problem-solving framework are represented by indicators 1 to 9. Scientific inquiry emphasizes understanding a problem, generating hypotheses, and seeking and developing reasonable arguments. Therefore, science problem-solving participates in the integrated STEM-based mathematical problem-solving framework through indicators 1 to 5 and 7 to 9. Engineering problem-solving emphasizes creating (designing) prototypes, building models, and testing solutions. Testing solutions is defined as evaluating and ensuring the correctness of the answer. Hence, the steps representing engineering and technology problem-solving in integrated STEM-based mathematical problem-solving are indicators 1, 2, and 5 to 9. Technology focuses on applying tools during problem-solving tasks (e.g., calculator, ruler, grid paper, etc.). Some indicators are shared across disciplines (e.g., indicator 5), while other indicators are more domain-specific (e.g., indicator 3). The framework for integrated STEM-based mathematical problem solving is described in Table 1.1.

The problem-solving indicators are ordered into a flowchart to demonstrate that problem-solving is a process. However, the indicators may occur repeatedly and in any sequence. The evaluation, reflection, and conclusion indicators can occur after any problem-solving indicator. This is because an evaluation process is necessary to address a conclusion before making a decision. The indicator related to representing a problem occurs in parallel with the step of identifying given and unknown information. Figure 1.2 illustrates the flowchart of problem-solving indicators.

Table 1. 1 An Integrated STEM-Based Mathematical Problem-Solving Framework

Phases	Indicators
Exploring and understanding	1. Determining the goal of a problem (STEM); 2. Identifying relevant and valuable unknowns and given information (STEM); 3. Employing a useful basic concept (SM)
Representing and formulating	4. Constructing a tabular, graphical, symbolic, or verbal representation of a problem using technology (STM); 5. Making a prediction (hypothesis), developing criteria, or checking existing theory (STEM).
Planning and executing	6. Formulating a reasonable argument, explanation, and solution (including strategy, step design, and model building) (TEM); 7. Arranging, critically choosing, and evaluating alternatives by deducting, proofing, or finding a counterexample (STEM); 8. Performing a plan by applying mathematics concepts, mathematization, reasoning, computational skills, science concepts, and technology (STEM).
Monitoring and reflecting	9. Drawing conclusions, evaluating, and reflecting on results and methods (STEM).

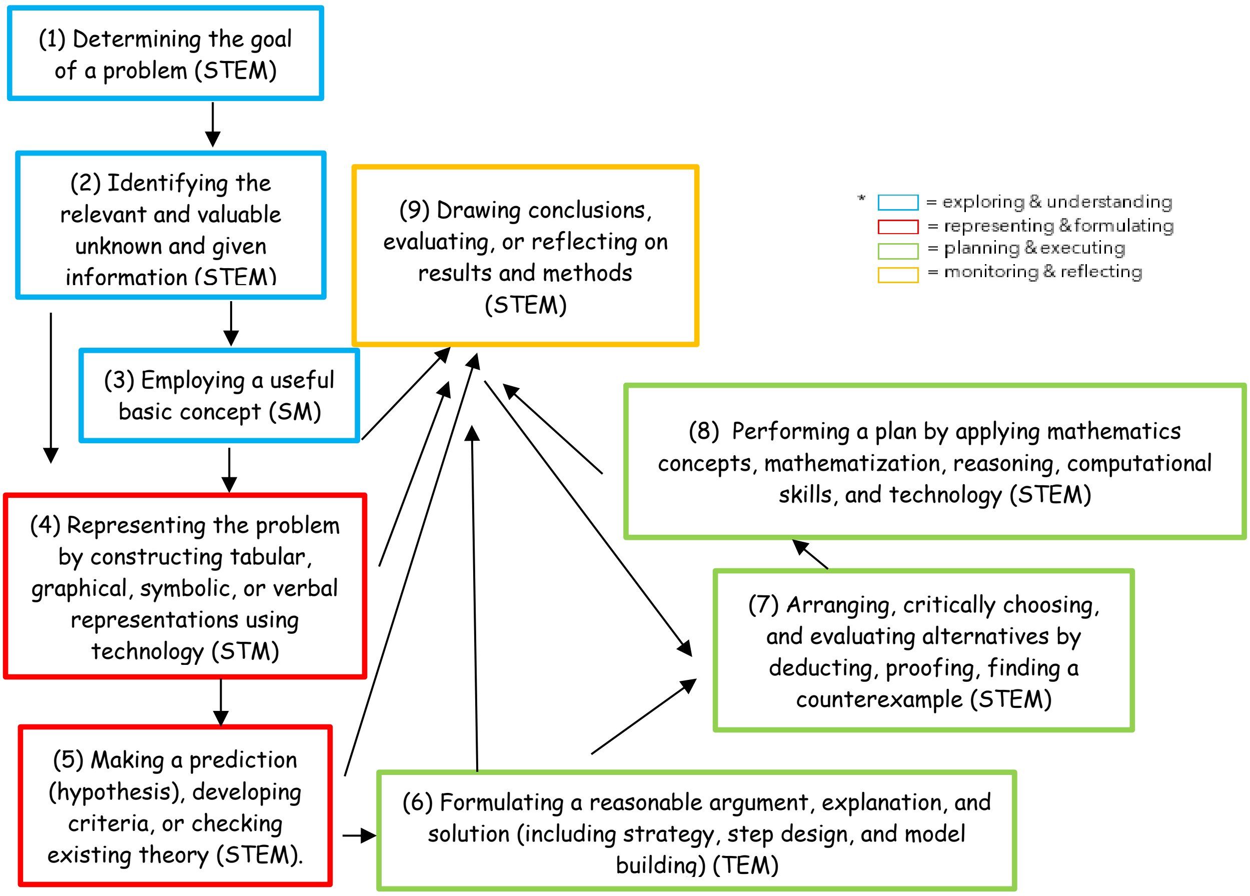


Figure 1. 2 Visual Representation of the Integrated STEM-Based Mathematical Problem-Solving Framework

1.4 The Present Dissertation

The present dissertation uses a study-based format. The main aim of the present dissertation is to assess middle school students' mathematical problem-solving skills and the factors influencing them. To achieve this aim, a review study and four empirical studies were conducted with the following objectives: (1) review relevant assessments and frameworks in STEM problem-solving; (2) develop and validate an integrated STEM-based mathematical problem-solving test for assessing mathematical problem-solving skills; (3) develop and validate a mathematics DSPK test to measure prior mathematical knowledge; (4) investigate the profiles, development, and differences of mathematical problem-solving skills based on students' background variables; and (5) examine the cognitive and SES factors influencing mathematical problem-solving skills. Hence, the general research questions of this dissertation are listed below.

1. What is the trend of STEM problem-solving assessments, including instruments, topics, participants, years, disciplines, and frameworks used?
2. Is the integrated STEM-based mathematical problem-solving test we have developed suitable for the assessment of mathematical problem-solving?
3. Is the mathematics DSPK test we have developed suitable for the assessment of prior mathematical knowledge?
4. What are the profiles, development, and differences of students' mathematical problem-solving skills based on grades, gender, and school locations?
5. To what extent do mathematics DSPK, science knowledge, text comprehension skills, parents' education, and family income variables interact in predicting students' mathematical problem-solving skills?

The first research question reviews the current trends in STEM problem-solving, focusing on instruments, disciplines, years, topics, participants, and frameworks. This was done to determine the appropriate participant, assessment tool, and framework for the current study. Previous study, such as Gao et al. (2020), has provided a general overview of STEM assessment, but often lacks detailed categorization of instruments specifically targeting problem-solving skills. Furthermore, they did not evaluate the advantages and limitations of existing tools in depth. Therefore, this study narrows the focus to STEM problem-solving skills and provides detailed drawbacks and advantages of the instruments. STEM problem-solving was chosen for review rather than only mathematical problem-solving because an integrated STEM-based mathematical problem-solving test will be developed. By examining STEM problem-solving more broadly, the frameworks and contextual features used across disciplines can be identified. This allows to incorporate relevant elements into the design of a new and comprehensive framework for an integrated STEM-based mathematical problem-solving test.

The next section consists of four empirical studies. The research question of the first empirical study is to develop and validate an integrated STEM-based mathematical problem-solving test for assessing mathematical problem-solving skills. This research is grounded in prior works on interdisciplinary STEM task development, which emphasize the importance of assessing students' ability to solve real-world and cross-disciplinary problems (Scherer & Tiemann, 2012; Shanta, 2019; Shute et al., 2016). However, existing instruments often underrepresent mathematical aspects and are difficult to implement in classroom settings. Therefore, the integrated STEM-based mathematical problem-solving test in this study employs a scenario-based essay format situated in an interdisciplinary STEM context, with a strong emphasis on mathematics. It is specially designed to be classroom-friendly and to assess the process of problem-solving, particularly within the Indonesian context.

The research question of the second empirical study aims to develop and validate the mathematics DSPK test for assessing prior mathematical knowledge that affects mathematical problem-solving skills. This research builds on previous studies that have developed tests of prior knowledge (Al-Mutawah et al., 2019; Binder et al., 2019; Hailikari et al., 2007). However, many of these instruments are either oriented toward science-based readiness or focus narrowly on specific mathematical topics. Therefore, the mathematics DSPK test in this study is designed to cover several core mathematics topics relevant to the Indonesian curriculum. Furthermore, this also addresses the need for a prior mathematical knowledge test that is aligned with and supports the integrated STEM-based mathematical problem-solving test.

The third empirical study is a cross-sectional study to examine the profiles, development, and differences of students' mathematical problem-solving skills based on their grades, gender, and school locations. This research question is grounded in previous studies that examined differences in mathematics achievement across demographic variables (Anjum, 2015; Nepal, 2017; Ramos et al., 2021), investigated the development of problem-solving skills (Molnár et al., 2013), and explored the overall profiles of students' mathematical problem-solving skills (Novriani & Surya, 2017). However, these earlier studies did not conduct detailed analyses of these differences and were limited to monodisciplinary tasks. Therefore, this study provides an in-depth investigation of the subsamples contributing to differences in problem-solving skills using an integrated STEM-based mathematical problem-solving task. It also explores the detailed development across each phase of problem-solving, aiming to address limitations found in previous similar studies.

The last empirical study examines the factors influencing mathematical problem-solving skills. It emphasizes cognitive (mathematics DSPK, science knowledge, and text comprehension skills) and SES (family income and parents' education) factors. This research question is grounded in previous studies that identified these factors as significant predictors of mathematical problem-solving skills, although typically using monodisciplinary tasks and analyzing the factors separately (Alghazo & Alghazo, 2015; Bahar, 2013; Daroczy et al., 2020; Marks & Pokropek, 2019; Scherer & Beckmann, 2014; Vilenius-Tuohimaa et al., 2008). Therefore, in this study, these factors are combined into a new theoretical model and analyzed using SEM. It will give comprehensive direct and indirect relationships among these factors and mathematical problem-solving skills that are missing in previous studies. Additionally, this study compares the results of several developed models to obtain more accurate and convincing results. Furthermore, this study uses an integrated STEM-based mathematical problem-solving task to measure mathematical problem-solving skills.

1.5 The Structure of the Dissertation

The dissertation is organized into three chapters including a general introduction, review and empirical studies, and discussion, conclusions, and educational implications. Chapter 1 highlights the background and rationale of the study, including the gaps, significance, and essence. It also reviews both theoretical and Indonesian policy-related literature on mathematical problem-solving skills and explains the factors that influence them. Additionally, the chapter also introduces a theoretical framework guiding the study, outlines the aims of the dissertation, describes the structure of the dissertation, and provides the general methodology.

Chapter 2 addresses the review and empirical studies. Study 1 is a systematic review study that describes the trends and main characteristics (including frameworks, content of assessment, etc.) of assessment in mathematical problem-solving and each STEM discipline. It assesses through a systematic review of 77 articles in 2010 to 2020. It also provides a detailed discussion of examples of assessments in STEM disciplines. These findings have been published in *the International Journal of Assessment and Evaluation* (Amalina & Vidákovich, 2022b).

The empirical studies are also presented in Chapter 2, incorporating studies 2 to 5. Studies 2 and 3 discuss the development and validation of the integrated STEM-based mathematical problem-solving test and the mathematics DSPK test. Two different cross-sectional studies with different sample sizes of middle school students were conducted to validate each instrument. The instruments were validated for both construct and content. Content validation was conducted by practitioners, while the construct validation was analyzed using the Rasch model measurement. In addition, it discusses the revision based on the results.

The work of developing and reporting psychometric evidence of the integrated STEM-based mathematical problem-solving test has been published in the *Journal on Mathematics Education* (Amalina & Vidákovich, 2022a) and was presented at the *XXII Országos Neveléstudományi Konferencia* (Amalina & Vidákovich, 2022d) and the *JURE* conference (Amalina & Vidákovich, 2023d). In addition, the result of the development and validation of the mathematics DSPK test has been published in the *International Journal of Evaluation and Research in Education (IJERE)* (Amalina & Vidákovich, 2023a) and was presented at the *JURE* conference (Amalina & Vidákovich, 2022c). Although the publication of Study 3 uses the term "mathematical problem-solving test" in its title, this dissertation clearly differentiates between two uses of the term. In Study 3, it refers specifically to the mathematics DSPK test, which assesses students prior conceptual and procedural knowledge in the steps involved in problem-solving. In contrast, the term "mathematical problem-solving" in Studies 2, 4, and 5 refers to the integrated STEM-based mathematical problem-solving assessment.

Study 4 is an empirical study on the profiles and development of mathematical problem-solving skills, and differences based on grades, gender, and school locations. The results are presented in pirate graphs to detect the development across grade cohorts, performed by R Studio. Differences in mathematical problem-solving skills were analyzed using one-way ANOVA and a t-test. The findings of students' development and differences in mathematical problem-solving skills have been published in the *Heliyon* (Amalina & Vidákovich, 2023c).

Study 5 focuses on the cognitive and SES factors influencing mathematics problem-solving skills. The cognitive factor includes mathematics DSPK, science knowledge, and text comprehension skills. The SES factor covers father's and mother's educations and family income. The results inform the factors that predicted mathematical problem-solving skills the most. The finding of factors influencing mathematics problem-solving has been published in the *Heliyon* (Amalina & Vidákovich, 2023b) and was presented at the *JURE* conference (Amalina & Vidákovich, 2024).

Finally, the last chapter addresses the general discussion and conclusions that interpret and compare the results with previous studies. In addition, the educational implications, limitations, and future directions are emphasized.

1.6 General Methodology of the Empirical Studies

1.6.1 Participants

The population of the study is middle school students (grades 7 to 9) from A-accreditation schools in both rural and urban areas in East Java, Indonesia. The present study specified the participants because the target skill assessed is higher-order thinking skills. A-accreditation schools have stable quality. Additionally, East Java is one of the largest provinces that has a well-constructed educational system. Hence, students are familiar with higher-order thinking skills. There are 1,886 middle schools with A-accreditation in East Java, Indonesia, with a total of approximately 2 million students (npd.kemendikbud.go.id). A total of 1,067 students participated in the main study, ensuring a 95% confidence level with a 3% margin of error.

There are three different sample sizes in this dissertation. In the pilot study (Studies 2 and 3) to report the psychometric properties of the developed tests, the convenience sampling method was used. In addition, in the main study (Studies 4 and 5) to investigate the development, differences, and factors influencing mathematical problem-solving skills, a stratified random sampling method was applied. First, schools with A-accreditation were selected. Then, classes were randomly chosen from within those schools (see Table 1.2 for the number of participants in each study).

1.6.2 Instruments

The main instrument of the study is an integrated STEM-based mathematical problem-solving test for grades seven to nine. The test measures students' mathematical problem-solving skills using a STEM context. It is an essay-based mathematical scenario problem related to scientific phenomena that was developed under the integrated STEM-based mathematical problem-solving framework and the 2013 Indonesian curriculum. The test utilizes engineering-based design and scientific inquiry in the problem-solving process by using technology as a tool (e.g., calculator, grid paper, etc.). STEM serves as both the context and methodology to solve the problem, while mathematics plays a major role in the content of the test. There are six scenarios, with three scenarios for each grade, including one to two anchor scenarios. The test assesses topics such as numbers and measurement, ratios and proportions, geometry, and statistics. Each scenario includes eight metacognitive prompt items to scaffold students in exploring their problem-solving skills. The maximum score for each prompting item is five. The test has been proven robust to measure integrated STEM-based mathematical problem-solving skills.

The background questionnaire is included at the end of the integrated-STEM based mathematical problem-solving test. Its purpose is to gather information on the SES (such as parents' education: from elementary school to doctoral levels, parents' job, family income: four levels based on the average income in the province, and the number of siblings) and demographic characteristics (such as gender, age, grade, ethnicity, and school background). Family income was grouped based on the total parental income relative to the number of dependent children. The SES data provided by students were confirmed using the SES data from the schools, which were obtained from parents. The background questionnaire was validated by the teachers before being administered.

The second instrument is the mathematics DSPK test, a preliminary test assessing students' prior mathematical knowledge. It includes basic conceptual and procedural knowledge relevant to the steps involved in problem-solving. It was developed according to the 2013 Indonesian mathematics curriculum. The test is a single test for all grades that includes 30 multiple choice questions with four options. The topics align with those used in the integrated STEM-based mathematical problem-solving test and the anchor topics for each grade. The maximum score is 30. The test has demonstrated good psychometric properties.

Another instrument to measure text comprehension skills is the diagnostic test of the Indonesian language literacy test 2022, provided by the Indonesian MOE. The test is a single test for grades seven to nine that was conducted for 60 minutes at the beginning of the 2022/2023 academic year. The current study evaluates text comprehension skills using score data from this test. It assesses students' ability to find information, analyze and integrate it, and assess and reflect on scientific and linguistic texts. This test includes 21 items in matching type questions, short response questions, essays, and multiple-choice questions with simple and complex answers. Scores are displayed in percentages. The test has been shown to have acceptable construct validity and reliability.

The instrument to measure science knowledge is diagnostic multiple-choice tests related to environmental science topics (energy and photosynthesis, environmental issue and climate change, eco-friendly products, health and nutrition, biotechnology, and scientific procedure). The tests are provided in textbooks by the Indonesian MOE. Every school is required to use the same tests for each grade before learning each chapter. The mean scores from these diagnostic tests that provided by the schools were used to assess students' science knowledge in this study. Each test has nine items that have to be solved in 30 minutes. The test has been proven to have acceptable content validity and reliability.

1.6.3 General Procedures

Before conducting the main study, the development of the integrated STEM-based mathematical problem-solving and mathematics DSPK tests was performed. Subsequently, the developed instruments were analyzed for content and construct validity and reliability. The content validity was performed by practitioners, while the construct validity and reliability were analyzed using students' data from the pilot study. Data collection for the pilot study was carried out one day for each instrument in every school.

The researcher requested consent from the institutional review board of the Doctoral School of Education, University of Szeged, for the main study data collection. The researcher received an ethical approval letter in July 2022 with the reference number 7/2022. In addition, the researcher also provided a permission letter for conducting the study, which is signed by the researcher, the supervisor, and the head of the Institute of Education, University of Szeged. The ethical approval letter and the permission letters were given to the principals of A-accreditation schools. They were required to confirm their participation after a week. Schools that informed their participation facilitated the researcher to communicate with mathematics teachers to decide the data collection schedule. The data collection for the main study required two days in each class.

After the data collection process, the data were analyzed using descriptive and inferential statistical analyzes. Table 1.2 describes the general methodology of empirical studies.

Table 1. 2 The General Methodology of Empirical Studies

Aims	Timeline	Participant	Instruments & Procedures	Analyses
Validate a mathematics DSPK test.	December 2021	Grade 7 (n = 175)	Online DSPK test (Google form). Total time is 90 minutes.	Content validity index (CVI), intraclass correlation coefficient (ICC), Rasch analysis.
Validate an integrated STEM-based mathematical problem-solving test.	March-April 2022	Grades 7—9 (n = 286)	Paper-based integrated STEM-based mathematical problem-solving test and upload in google form. 3 hours with breaks in every hour.	
Investigate the development, profiles, and differences of the skills.	July-September 2022	Grades 7—9 (n = 1,067)	1st day: 3 hours with breaks in every hour (using an integrated STEM-based mathematical problem-solving test).	Descriptive statistics and compare means (one-way ANOVA and independent sample t-test).
Investigate factors influencing the skills.			2nd day: 90 minutes (using the DSPK test). Science tests and text comprehension test scores.	SEM (Path analysis).

1.7 References

Akbasli, S., Sahin, M., & Yaykiran, Z. (2016). The effect of reading comprehension on the performance in science and mathematics. *Journal of Education and Practice*, *7*(16), 108–121. https://files.eric.ed.gov/fulltext/EJ1108657.pdf

Al-Mutawah, M. A., Thomas, R., Eid, A., Mahmoud, E. Y., & Fateel, M. J. (2019). Conceptual understanding, procedural knowledge and problem-solving skills in mathematics: High school graduates work analysis and standpoints. *International Journal of Education and Practice*, *7*(3), 258–273. https://doi.org/10.18488/journal.61.2019.73.258.273

Alghazo, Y., & Alghazo, R. (2015). The effect of parental involvement and socioeconomic status on elementary students' mathematics achievement. *Journal of Social Sciences and Humanities*, *1*(5), 521–527. http://www.aiscience.org/journal/paperInfo/jssh?paperId=2312

Alomar, B. O. (2007). Personal and family factors as predictors of pupils' mathematics achievement. *Psychological Reports*, *101*(1), 259–269. https://doi.org/10.2466/PR0.101.1.259-269

Álvarez, J. A. M., Rhoads, K., & Campbell, R. C. (2019). Toward designing and developing Likert items to assess mathematical problem solving. In P. Liljedahl & M. Santos-Trigo (Eds.), *Mathematical problem solving* (pp. 231–260). Springer International Publishing. https://doi.org/10.1007/978-3-030-10472-6_11

Amalina, I. K., & Vidákovich, T. (2022a). 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
Amalina, I. K., & Vidákovich, T. (2022b). Assessment in STEM problem-solving: A systematic review. *The International Journal of Assessment and Evaluation*, *29*(2), 64–80. https://doi.org/https://doi.org/10.18848/2327-7920/CGP/v29i02/63-80
Amalina, I. K., & Vidákovich, T. (2023a). Assessment of domain-specific prior knowledge : A development and validation of mathematical problem-solving test. *International Journal of Evaluation and Research in Education (IJERE)*, *12*(1), 468–476. https://doi.org/10.11591/ijere.v12i1.23831
Amalina, I. K., & Vidákovich, T. (2023b). Cognitive and socioeconomic factors that influence the mathematical problem-solving skills of students. *Heliyon*, *9*(9), 1–14. https://doi.org/10.1016/j.heliyon.2023.e19539
Amalina, I. K., & Vidákovich, T. (2023c). Development and differences in mathematical problem-solving skills : A cross-sectional study of differences in demographic backgrounds. *Heliyon*, *9*(5), 1–11. https://doi.org/10.1016/j.heliyon.2023.e16366
Amalina, I. K., & Vidákovich, T. (2022c). A Validation of Mathematical Problem-Solving as A Domain-Specific Prior Knowledge Test. In JURE (Ed.), *JURE 2022-Unpredictable Challenges - Education in a Rapidly Changing World* (p. 45). JURE 2022. https://ssl.earli.org/sites/default/files/2022-07/JURE2022-BOA-1607.pdf
Amalina, I. K., & Vidákovich, T. (2023d). Psychometric Evidence of an Integrated STEM-based Mathematical Problem-Solving Test. In JURE (Ed.), *JURE 2023-Education as a hope in uncertain times* (p. 37). JURE 2023. https://www.earli.org/events/jure2023
Amalina, I. K., & Vidákovich, T. (2024). Structural Equation Modelling: Associating Cognitive Factors to Mathematical Problem-Solving Skills. In JURE (Ed.), *JURE 2024: Education and the Need for Dynamic Solutions: Shaping the Future of the Field* (p. 38). JURE 2024. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.earli.org/assets/files/JURE2024-BOA-180624.pdf
Amalina, I. K., & Vidákovich, T. (2022d). Validation of An Integrated STEM-based Mathematical Problem-solving Test. In J. Steklács & M.-K. Zsófia (Eds.), *XXII. Országos Neveléstudományi Konferencia- 21st century skills, literacy, equal opportunities* (p. 271). Hungarian Academy of Sciences Pedagogical Committee – PTE BTK Institute of Education.
Anjum, S. (2015). Gender difference in mathematics achievement and its relation with reading comprehension of children at upper primary stage. *Journal of Education and Practice*, *6*(16), 71–75. https://files.eric.ed.gov/fulltext/EJ1079951.pdf
Annaggar, A., & Tiemann, R. (2020). A video game as an assessment tool of problem-solving competence. *Gesellschaft Fur Didactik Der Chemie Und Physik (GDCP)*, *April*, 377–380. https://www.gdcp-ev.de/wp-content/tb2020/TB2020_377_Annaggar.pdf
Arieli, S., & Sagiv, L. (2018). Culture and problem-solving: Congruency between the cultural mindset of individualism versus collectivism and problem type. *Journal of Experimental Psychology: General*, *147*(6), 789–814. https://doi.org/10.1037/xge0000444
Baele, L. C. (2017). *Middle school engineering problem solving using traditional vs . E-PBL module instruction* (Issue January) [Aurora University].

https://search.proquest.com/openview/1cc284ab793aa6734c79e42d50a9a790/1?pq-origsite=gscholar&cbl=18750

Bahar, A. (2013). *The influence of cognitive abilities on mathematical problem solving performance* [The University of Arizona]. https://repository.arizona.edu/handle/10150/293594

Bayazit, I. (2013). An investigation of problem solving approaches, strategies, and models used by the 7th and 8th grade students when solving real-world problems. *Kuram ve Uygulamada Egitim Bilimleri*, *13*(3), 1920–1927. https://doi.org/10.12738/estp.2013.3.1419

Binder, T., Schmiemann, P., & Theyssen, H. (2019). Knowledge acquisition of biology and physics university students—the role of prior knowledge. Education Sciences, 9(4), 1–18. https://doi.org/10.3390/educsci9040281

Boero, P. (1999). Paolo Boero 1999. *International Newsletter on the Teaching and Learning of Mathematical Proof*. http://www.lettredelapreuve.org/OldPreuve/Newsletter/990708Theme/990708ThemeUK.html

Boonen, A. J. H., Van Wesel, F., Jolles, J., & Van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research*, *68*, 15–26. https://doi.org/10.1016/j.ijer.2014.08.001

Booth, J. L., & Davenport, J. L. (2013). The role of problem representation and feature knowledge in algebraic equation-solving. *Journal of Mathematical Behavior*, *32*(3), 415–423. https://doi.org/10.1016/j.jmathb.2013.04.003

Byrnes, J. P., & Takahira, S. (1994). Why some students perform well and others perform poorly on SAT-Math items. In *Contemporary Educational Psychology* (Vol. 19, Issue 1, pp. 63–78). https://doi.org/10.1006/ceps.1994.1007

Cano, F., García, Á., Berbén, A. B. G., & Justicia, F. (2014). Science learning: A path analysis of its links with reading comprehension, question-asking in class and science achievement. *International Journal of Science Education*, *36*(10), 1710–1732. https://doi.org/10.1080/09500693.2013.876678

Carifio, J. (2015). Updating, modernizing, and testing Polya's theory of [Mathematical] problem solving in terms of current cognitive, affective, and information processing theories of learning, emotions, and complex performances. *Journal of Education and Human Development*, *4*(3), 105–117. https://doi.org/10.15640/jehd.v4n3a12

Carlson, M. P., & Bloom, I. (2005). The cyclic nature of problem solving: An emergent multidimensional problem-solving framework. *Educational Studies in Mathematics*, *58*(1), 45–75. https://doi.org/10.1007/s10649-005-0808-x

Chamberlin, S. A. (2008). What is problem solving in the mathematics classroom? *Philosophy of Mathematics Education Journal*, *23*(1). http://socialsciences.exeter.ac.uk/education/research/centres/stem/publications/pmej/pome23/Chamberlin What is Math Prob Solving.doc

Cortina, K. S., Arel, S., & Smith-Darden, J. P. (2017). School Belonging in Different Cultures: The Effects of Individualism and Power Distance. *Frontiers in Education*, *2*(November), 1–12. https://doi.org/10.3389/feduc.2017.00056

Csapó, B., & Molnár, G. (2017). Potential for assessing dynamic problem-solving at the beginning of higher education studies. *Frontiers in Psychology*, *8*(NOV). https://doi.org/10.3389/fpsyg.2017.02022

Daroczy, G., Meurers, D., Heller, J., Wolska, M., & Nürk, H. C. (2020). The interaction of linguistic and arithmetic factors affects adult performance on arithmetic word problems. *Cognitive Processing*, *21*(1), 105–125. https://doi.org/10.1007/s10339-019-00948-5

Davis-Kean, P. E., Tighe, L. A., & Waters, N. E. (2021). The role of parent educational attainment in parenting and children's development. *Current Directions in Psychological Science*, *30*(2), 186–192. https://doi.org/10.1177/0963721421993116

Di Martino, P., & Signorini, G. (2019). Beyond the standardized assessment of mathematical problem solving competencies: From products to processes. In P. Liljedahl & M. Santos-Trigo (Eds.), *Mathematical problem solving* (pp. 209–229). Springer International Publishing. https://doi.org/10.1007/978-3-030-10472-6_10

Du Toit, S., & Du Toit, G. (2013). Learner metacognition and mathematics achievement during problem-solving in a mathematics classroom. *The Journal for Transdisciplinary Research in Southern Africa*, *9*(3), 505–518. https://doi.org/10.4102/td.v9i3.194

Engin-Demir, C. (2009). Factors influencing the academic achievement of the Turkish urban poor. *International Journal of Educational Development*, *29*(1), 17–29. https://doi.org/10.1016/j.ijedudev.2008.03.003

Fuentes, P. (1998). Reading comprehension in mathematics. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *72*(2), 81–88. https://doi.org/10.1080/00098659809599602

Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education*, *7*(1). https://doi.org/10.1186/s40594-020-00225-4

Gray, G. L., Costanzo, F., & Plesha, M. E. (2005). Problem solving in statics and dynamics: A proposal for a structured approach. *ASEE Annual Conference and Exposition, Conference Proceedings*, 11543–11554. https://doi.org/10.18260/1-2--15371

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts-something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, *105*(2), 364–379. https://doi.org/10.1037/a0031856

Grissom, J. B. (2004). Age and achievement. *Educational Policy Analysis Archives*, *12*(49). https://doi.org/https://doi.org/10.14507/epaa.v12n49.2004

Hailikari, T. (2009). *Assessing university students' prior knowledge* [University of Helsinki]. https://helda.helsinki.fi/bitstream/handle/10138/19841/assessin.pdf?sequence=1

Hailikari, T., Nevgi, A., & Lindblom-Ylänne, S. (2007). Exploring alternative ways of assessing prior knowledge, its components and their relation to student achievement: A mathematics based case study. *Studies in Educational Evaluation*, *33*(3–4), 320–337. https://doi.org/10.1016/j.stueduc.2007.07.007

Halpern, D. F., Wai, J., & Saw, A. (2005). A psychobiosocial model: Why females are sometimes greater than and sometimes less than males in math achievement. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 48–72). Cambridge University Press.

Heuvel-Panhuizen, M. Van Den. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics*, *25*(2), 2–9. https://www.jstor.org/stable/40248489

Imam, A., & Singh, G. P. (2015). Influence of gender, parental education, and parental occupation on mathematics achievement of secondary school students. *Indian Journal Od Research*, *4*(11), 187–190. https://www.worldwidejournals.com/paripex/recent_issues_pdf/2015/November/November_2015_1448104482__63.pdf

İncebacak, B. B., & Ersoy, E. (2016). Problem solving skills of secondary school students. *China-USA Business Review*, *15*(6), 275–285. https://doi.org/10.17265/1537-1514/2016.06.002

Attachment of Government Regulation Number 21 Year of 2016, 1 (2016).

Government Regulation Number 24 Year of 2016 Attachment 15, 1 (2016).

Government Regulation Number 24 Year of 2016 Attachment 6, 1 (2016).

Jäder, J., Lithner, J., & Sidenvall, J. (2020). Mathematical problem solving in textbooks from twelve countries. *International Journal of Mathematical Education in Science and Technology*, *51*(7), 1120–1136. https://doi.org/10.1080/0020739X.2019.1656826

Jitendra, A. K., Lein, A. E., Star, J. R., & Dupuis, D. N. (2013). The contribution of domain-specific knowledge in predicting students' proportional word problem-solving performance. *Educational Research and Evaluation*, *19*(8), 700–716. https://doi.org/10.1080/13803611.2013.845107

Jolly, A. (2016). *STEM by design: Strategies and activities for grades 4-8*. Routledge. https://doi.org/10.4324/9781315679976

Jones, I., Swan, M., & Pollitt, A. (2015). Assessing mathematical problem solving using comparative judgement. *International Journal of Science and Mathematics Education*, *13*(1), 151–177. https://doi.org/10.1007/s10763-013-9497-6

Junpeng, P., Marwiang, M., Chiajunthuk, S., Suwannatrai, P., Chanayota, K., Pongboriboon, K., Tang, K. N., & Wilson, M. (2020). Validation of a digital tool for diagnosing mathematical proficiency. *International Journal of Evaluation and Research in Education*, *9*(3), 665–674. https://doi.org/10.11591/ijere.v9i3.20503

Jurdak, M. (2016). STEM education as a context for real-world problem solving. In *Learning and teaching real world problem solving in school mathematics: A multiple-perspective framework for crossing the boundary* (pp. 151–163). Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-08204-2

Karatas, I., & Baki, A. (2013). The effect of learning environments based on problem solving on students ' achievements of problem solving. *International Electronic Journal of Elementary Education*, *5*(3), 249–267. https://www.iejee.com/index.php/IEJEE/article/view/25

Ke, F., & M. Clark, K. (2020). Game-based multimodal representations and mathematical problem solving. *International Journal of Science and Mathematics Education*, *18*(1), 103–122. https://doi.org/10.1007/s10763-018-9938-3

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1). https://doi.org/10.1186/s40594-016-0046-z

Khairani, A. Z., & Sahari Nordin, M. (2011). The development and construct validation of the mathematics proficiency test for 14-year-old students. *Asia Pacific Journal of Educators and Education*, *26*(1), 33–50. http://web.usm.my/apjee/webtest/APJEE_26.1.2011/26.1.2011_33-50.pdf

Kodippili, A. (2011). Parents' education level in students' mathematics achievement; do school factors matter? *Academic Leadership*, *9*(1). https://scholars.fhsu.edu/alj/vol9/iss1/39/

Lasa, A., Abaurrea, J., & Iribas, H. (2020). Mathematical content on STEM activities. *Journal on Mathematics Education*, *11*(3), 333–346. https://doi.org/10.22342/JME.11.3.11327.333-346

Lee, C. Y., & Chen, M. P. (2009). A computer game as a context for non-routine mathematical problem solving: The effects of type of question prompt and level of prior knowledge. *Computers and Education*, *52*(3), 530–542. https://doi.org/10.1016/j.compedu.2008.10.008

Lee, J., & McIntire, W. G. (2000). Interstate variation in the mathematics achievement of rural and nonrural students. *Journal of Research in Rural Education*, *16*(3), 168–181. https://jrre.psu.edu/sites/default/files/2019-08/16-3_2.pdf

Lester, F. K. (2013). Thoughts about research on mathematical problem- solving. *The Mathematics Enthousiasm*, *10*(1), 245–278. https://doi.org/https://doi.org/10.54870/1551-3440.1267

Leung, F. K. S. (2001). In Search of an East Asian Identity in Mathematics Education. *Educational Studies in Mathematics*, *47*, 35–51. https://doi.org/10.1007/978-94-010-9046-9_34

Liljedahl, Peter, Santos-Trigo, M., Malaspina, U., & Bruder, R. (2016). Problem solving in mathematics education. In G. Kaiser (Ed.), *ICME-13 Topical Surveys*. Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_129

Lin, X. (2021). Investigating the unique predictors of word-problem solving using meta-analytic structural equation modeling. *Educational Psychology Review*, *33*(3), 1097–1124. https://doi.org/10.1007/s10648-020-09554-w

Maass, K., Geiger, V., Ariza, M. R., & Goos, M. (2019). The role of mathematics in interdisciplinary STEM education. *ZDM - Mathematics Education*, *51*(6), 869–884. https://doi.org/10.1007/s11858-019-01100-5

Männamaa, M., Kikas, E., Peets, K., & Palu, A. (2012). Cognitive correlates of math skills in third-grade students. *Educational Psychology*, *32*(1), 21–44. https://doi.org/10.1080/01443410.2011.621713

Marks, G. N., & Pokropek, A. (2019). Family income effects on mathematics achievement: their relative magnitude and causal pathways. *Oxford Review of Education*, *45*(6), 769–785. https://doi.org/10.1080/03054985.2019.1620717

Meng, J., & Liu, S. (2022). Effects of Culture on the Balance Between Mathematics Achievement and Subjective Wellbeing. *Frontiers in Psychology*, *13*(May), 1–11. https://doi.org/10.3389/fpsyg.2022.894774

Miteva, M., Zlatanovska, B., & Kocaleva, M. (2022). Some examples for teaching mathematics in STEM context. *STEM Education Notes*, *1*(2), 31–39. https://doi.org/10.37418/stem.1.2.1

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, *9*, 35–45. https://doi.org/10.1016/j.tsc.2013.03.002

Nepal, B. (2017). Impact of gender and location on mathematical thinking and mathematics achievement. *Journal of Advanced Academic Research*, *3*(3), 11–21. https://doi.org/10.3126/jaar.v3i3.16803

Niu, J., Xu, H., & Yu, J. (2025). Identifying multilevel factors on student mathematics performance for Singapore, Korea, Finland, and Denmark in PISA 2022: considering individualistic versus collectivistic cultures. *Humanities and Social Sciences Communications*, *12*(1), 1–12. https://doi.org/10.1057/s41599-025-04466-y

Novriani, M. R., & Surya, E. (2017). Analysis of Student difficulties in mathematics problem solving ability at MTS Swasta IRA Medan. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, *33*(3), 63–75. https://core.ac.uk/download/pdf/249335829.pdf

Nunokawa, K. (2005). Mathematical problem solving and learning mathematics: What we expect students to obtain. *Journal of Mathematical Behavior*, *24*(3–4), 325–340. https://doi.org/10.1016/j.jmathb.2005.09.002

OECD. (2019). Programme for International Student Assessment (PISA). Results From PISA 2018: Indonesia. *OECD Education Working Papers*. https://doi.org/10.1007/978-94-6209-497-0_69

Pangeni, K. P. (2014). Factors determining educational quality: Student mathematics achievement in Nepal. *International Journal of Educational Development*, *34*(1), 30–41. https://doi.org/10.1016/j.ijedudev.2013.03.001

Polya, G. (1945). *How to solve it*. Princeton University.

Priemer, B., Eilerts, K., Filler, A., Pinkwart, N., Rösken-Winter, B., Tiemann, R., & Zu Belzen, A. U. (2020). A framework to foster problem-solving in STEM and computing education. *Research in Science and Technological Education*, *38*(1), 105–130. https://doi.org/10.1080/02635143.2019.1600490

Ramos, R., Duque, J. C. C., & Nieto, S. (2021). Decomposing the rural-urban differential in student achievement in Colombia using PISA microdata. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2051358

Robinson, N. (2016). *A case study exploring the effects of using an integrative STEM curriculum on eighth grade students ' performance and engagement in the mathematics classroom*. Georgia State University.

Salgado, A. J. F. (2017). The role of context and context familiarity on mathematics problems. *Revista Latinoamericana de Investigacion En Matematica Educativa*, *20*(3), 265–292. https://doi.org/10.12802/relime.17.2031

Scherer, R., & Beckmann, J. F. (2014). The acquisition of problem solving competence: evidence from 41 countries that math and science education matters. *Large-Scale Assessments in Education*, *2*(1), 1–22. https://doi.org/10.1186/s40536-014-0010-7

Scherer, R., & Tiemann, R. (2012). Factors of problem-solving competency in a virtual chemistry environment: The role of metacognitive knowledge about strategies. *Computers and Education*, *59*(4), 1199–1214. https://doi.org/10.1016/j.compedu.2012.05.020

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press, Inc.
Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast*, *10*(1), 9–34. https://scholarworks.umt.edu/cgi/viewcontent.cgi?article=1258&context=tme
Serafino, K., & Cicchelli, T. (2003). Cognitive theories, prior knowledge, and anchored instruction on mathematical problem solving and transfer. *Education and Urban Society*, *36*(1), 79–93. https://doi.org/10.1177/0013124503257016
Shanta, S. (2019). Developing and assessing authentic problem-solving skills in high school pre-engineering students. *ASEE Annual Conference and Exposition, Conference Proceedings*. https://doi.org/10.18260/1-2--32637
Shaughnessy, M. (2013). Mathematics in a STEM context. *National Council of Teachers of Mathematics*, *18*(6), 324. http://www.jstor.org/stable/10.5951/mathteacmiddscho.18.6.0324
Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106–117. https://doi.org/10.1016/j.chb.2016.05.047
Spencer, M., Fuchs, L. S., & Fuchs, D. (2020). Language-related longitudinal predictors of arithmetic word problem solving: A structural equation modeling approach. *Contemporary Educational Psychology*, *60*(November 2019), 101825. https://doi.org/10.1016/j.cedpsych.2019.101825
Suratno, Wahono, B., Chang, C. Y., Retnowati, A., & Yushardi, Y. (2020). Exploring a direct relationship between students' problem-solving abilities and academic achievement: A STEM education at a coffee plantation area. *Journal of Turkish Science Education*, *17*(2), 211–224. https://doi.org/10.36681/tused.2020.22
Süß, H. M., & Kretzschmar, A. (2018). Impact of cognitive abilities and prior knowledge on complex problem solving performance - Empirical results and a plea for ecologically valid microworlds. *Frontiers in Psychology*, *9*(MAY), 1–22. https://doi.org/10.3389/fpsyg.2018.00626
Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, *96*(3), 471–491. https://doi.org/10.1037/0022-0663.96.3.471
Timm, M., & Uibu, K. (2015). Development of student text comprehension and language semantics in primary school. *Procedia - Social and Behavioral Sciences*, *191*, 793–800. https://doi.org/10.1016/j.sbspro.2015.04.471
Tuong, H. A., Nam, P. S., Hau, N. H., Lavicza, Z., Hougton, T., Thi, V., & Tien, B. (2023). Utilising STEM-based practices to enhance mathematics teaching in Vietnam: Developing students' real-world problem-solving and 21st century. *Journal of Technology and Science Education*, *13*(1), 79–91. https://doi.org/https://doi.org/10.3926/jotse.1790
Vilenius-Tuohimaa, P. M., Aunola, K., & Nurmi, J. E. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology*, *28*(4), 409–426. https://doi.org/10.1080/01443410701708228
Vista, A. (2013). The role of reading comprehension in maths achievement growth:

Investigating the magnitude and mechanism of the mediating effect on maths achievement in australian classrooms. *International Journal of Educational Research*, *62*, 21–35. https://doi.org/10.1016/j.ijer.2013.06.009

Vukelic, A., Jakšetić, J., Praljak, M., Krnić, M., Burić, T., Himmelreič, K. K., Andrić, M., Lovričević, N., Perić, J., Özçağ, S., Özçağ, E., Yıldız, F., Dost, Ş., Sadiku, M., Snopçe, H., Radevski, V., Tuneski, N., Koceva-Lazarova, L., Miteva, M., … Jolevska-Tuneska, B. (2020). *The guideline of mathSTEM method: Teaching mathematics in STEM context for STEM students*. Erasmus +, Agency for Mobility and EU Programmes, mathSTEM.

Wells, J. G. (2016). PIRPOSAL model of integrative STEM education: Conceptual and pedagogical framework for classroom implementation. *Technology and Engineering Teacher*, *75*(6), 12–19. http://america.aljazeera.com/articles/2013/9/16/many-us-bridges-arestructurallyunsoundsaysnewreport.html

Widodo, S. A., Ibrahim, I., Hidayat, W., Maarif, S., & Sulistyowati, F. (2021). Development of mathematical problem solving tests on geometry for junior high school students. *Jurnal Elemen*, *7*(1), 221–231. https://doi.org/10.29408/jel.v7i1.2973

Williams, J. H. (2005). Cross-national variations in rural mathematics achievement: A descriptive overview. *Journal of Research in Rural Education*, *20*(5), 1–18. https://jrre.psu.edu/sites/default/files/2019-08/20-5.pdf

Wilson, J. W., Fernandez, M. L., & Hadaway, N. (1989). Mathematical problem solving. *The American Mathematical Monthly*, *96*(1), 68. https://doi.org/10.2307/2323271

Woods, D. R. (2000). An evidence-based strategy for problem solving. *Journal of Engineering Education*, *89*(4), 443–459. https://doi.org/10.1002/j.2168-9830.2000.tb00551.x

Wu, M., & Adams, R. J. (2006). Modelling mathematics problem solving item responses using a multidimensional IRT model. *Mathematics Education Research Journal*, *18*(2), 93–113. https://doi.org/10.1007/BF03217438

Zeeshan, K., Watanabe, C., & Neittaanmaki, P. (2021). Problem-solving skill development through STEM learning approaches. *Proceedings - Frontiers in Education Conference, FIE*, *2021-Octob*. https://doi.org/10.1109/FIE49875.2021.9637226

Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A review of literature. *International Education Journal*, *8*(2), 187–203. https://files.eric.ed.gov/fulltext/EJ834219.pdf

Study 1: Assessment in STEM problem-solving: a systematic review.

(Amalina & Vidákovich, 2022b)

VOLUME 29 ISSUE 2

The International Journal of
Assessment and Evaluation

Assessment in STEM Problem-Solving

A Systematic Review

IJTIHADI KAMILIA AMALINA AND TIBOR VIDÁKOVICH

COMMON GROUND

THELEARNER.COM

THE INTERNATIONAL JOURNAL OF ASSESSMENT AND EVALUATION
https://thelearner.com
ISSN: 2327-7920 (Print)
ISSN: 2327-8692 (Online)
https://doi.org/10.18848/2327-7920/CGP (Journal)

First published by Common Ground Research Networks in 2022
University of Illinois Research Park
60 Hazelwood Drive
Champaign, IL 61820 USA
Ph: +1-217-328-0405
https://cgnetworks.org

The International Journal of Assessment and Evaluation is a peer-reviewed, scholarly journal.

Common Ground Research Networks, a member of Crossref

PEER REVIEW
Articles published in *The International Journal of Assessment and Evaluation* are peer reviewed using a two-way anonymous peer review model. Reviewers are active participants of The Learner Research Network or a thematically related Research Network. The publisher, editors, reviewers, and authors all agree upon the following standards of expected ethical behavior, which are based on the Committee on Publication Ethics (COPE) Core Practices. More information can be found at https://cgnetworks.org/journals/publication-ethics.

ARTICLE SUBMISSION
The International Journal of Assessment and Evaluation publishes biannually (June, December). To find out more about the submission process, please visit https://thelearner.com/journals/call-for-papers.

ABSTRACTING AND INDEXING
For a full list of databases in which this journal is indexed, please visit https://thelearner.com/journals/collection.

RESEARCH NETWORK MEMBERSHIP
Authors in *The International Journal of Assessment and Evaluation* are members of The Learner Research Network or a thematically related Research Network. Members receive access to journal content. To find out more, visit https://thelearner.com/about/become-a-member.

SUBSCRIPTIONS
The International Journal of Assessment and Evaluation is available in electronic and print formats. Subscribe to gain access to content from the current year and the entire backlist. Contact us at cgscholar.com/cg_support.

ORDERING
Single articles and issues are available from the journal bookstore at https://cgscholar.com/bookstore.

OPEN RESEARCH
The International Journal of Assessment and Evaluation is Hybrid Open Access, meaning authors can choose to make their articles open access. This allows their work to reach an even wider audience, broadening the dissemination of their research. To find out more, please visit https://cgnetworks.org/journals/open-research.

Assessment in STEM Problem-Solving: A Systematic Review

Ijtihadi Kamilia Amalina,[1] University of Szeged, Hungary
Tibor Vidákovich, University of Szeged, Hungary

Abstract: Problem-solving related to science, technology, engineering, and mathematics (STEM) is a core criterion in education in dealing with the challenges of the twenty-first century. Success in STEM problem-solving can be measured through assessment. However, lack of information regarding a review of assessment in STEM problem-solving is a major problem. This study aims to provide a systematic overview of the trend assessment used and detailed information related to topics, participants, and frameworks in STEM problem-solving. A Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) model was used to extract articles published between 2010 and 2020 concerning inclusion and exclusion criteria. The selected articles are analyzed in relation to bias regarding risk and categorized based on disciplines and learning domains. A total of seventy-seven articles were organized according to these criteria, mostly in the science and mathematics areas conducted in middle school students. It was found that the most common assessments in monodisciplinary and interdisciplinary areas were an essay test (n = 34) and a complex problem scenario test (n = 11), respectively. Program for International Student Assessment (PISA) content was widely used in STEM problems. Engineering design and technology served as an integrator in cognitive interdisciplinary and transdisciplinary STEM problem-solving assessments. Questionnaires and in-depth assessments were also used, primarily to measure metacognitive and affective factors in monodisciplinary and interdisciplinary areas. The issues of sample size, specification of assessment, framework, challenges, drawbacks, and advantages will be discussed. This study supports future research in suggesting the best assessment in STEM problem-solving.

Keywords: Assessment, Problem-Solving, STEM, Systematic Review

Introduction

Problem-solving is a key skill needed for twenty-first-century challenges and future jobs (Shute et al. 2016). Ongoing challenges in economic, scientific, engineering, technology, sociocultural, and general life skills require students to be good at analyzing problems and making decisions (Shanta 2019; Shanta and Wells 2020). For these reasons, problem-solving has been set as one of the core standards in the curriculum for students, particularly in mathematics, science, and engineering (Chang 2010; Karatas and Baki 2017).

It is necessary to embed several subjects in problem-solving to foster an interdisciplinary mindset (Özcan 2016; Shanta and Wells 2020). However, it is evident that students still encounter difficulties when solving real-world problems (Shanta and Wells 2020). This is because (1) such problems are learned and assessed separately with a focus on results rather than the process (Shanta 2019); (2) problem-solving in an academic setting is a well-defined and constrained challenge compared with a real-life situation (Schwortz and Burrows 2020; Shute et al. 2016); and (3) a researcher might not focus on other factors (e.g., affective and metacognition) that affect problem-solving (Guven and Cabakcor 2013; Soobard and Rannikmäe 2014).

Science, technology, engineering, and mathematics (STEM) education embodies more effective problem-solving activities that focus on integrating two or more subjects closely related to real life (Kelley and Knowles 2016). STEM prepares students to encounter global economic competitiveness and career challenges, as well as the jobs needed to support innovation and manufacturing (Robinson 2016; Shahali et al. 2017).

[1] Corresponding Author: Ijtihadi Kamilia Amalina, Doctoral School of Education, University of Szeged, Szeged, 6722, Hungary. email: ijtihadi.kamilia.amalina@edu.u-szeged.hu

The International Journal of Assessment and Evaluation
Volume 29, Issue 2, 2022, https://thelearner.com

ISSN: 2327-7920 (Print), ISSN: 2327-8692 (Online)
https://doi.org/10.18848/2327-7920/CGP/v29i02/63-80 (Article)

The issue of students' lack of interest in STEM courses contributes to lower enrollment in STEM majors, necessitating an assessment to track the reasons for such disinterest (Shahali et al. 2019). Moreover, assessments measuring student achievement can be drawn on to transfer knowledge from the teaching and learning process to solve problems (Eseryel et al. 2014; Shanta 2019). Hence, the development of valid assessments of the cognitive, affective, and psychomotor factors in problem-solving is a central goal in education (Eseryel, Ifenthaler, and Ge 2013; Scherer and Tiemann 2012; Shanta 2019).

Over the past 40 years, researchers have developed instruments to assess problem-solving, most of which focus only on the monodisciplinary area (e.g., Docktor et al. 2015; Chang 2010; Chusinkunawut et al. 2018). Currently, the assessments focus not only on summative but also on formative aspects. Project-based learning is a formative assessment that measures student performance in STEM problem-solving (Capraro and Corlu 2013; Kelley and Knowles 2016; Yu, Fan, and Lin 2015). Furthermore, the summative assessment that can be done before or after learning is acquired through problem-based assessment (Eseryel et al. 2014).

International and well-known tests in STEM problem-solving have been foundational to developing other assessments. The Program for International Student Assessment (PISA) embodies standard international assessments for problem-solving closely related to real-life contexts (OECD 2013). Moreover, a well-known computer-based diagnostic assessment developed by Greiff and Funke (2009), called MicroDYN, requires participants to detect causal relations in everyday activity. Researchers utilized the MicroDYN because of its suitability for measuring interdisciplinary traits (e.g., Annaggar and Tiemann 2020; Weise, Greiff, and Sparfeldt 2020).

Various types of assessment tools in STEM problem-solving have been developed by researchers. However, several worthwhile assessments still need to be explored for further development in STEM problem-solving (e.g., Chusinkunawut et al. 2018; Eseryel, Ifenthaler, and Ge 2013; Scherer and Tiemann 2012; Shanta 2019). The problem is the inadequate study of assessments in this area, highlighting the need for detailed evaluation of assessment frameworks and the best assessment for future study. The main parts related to STEM problem-solving assessment that need to be described involve both general and specific information. General information includes the distribution of articles over the year, the distribution of participants, and disciplinary distribution combined with the learning domain. The specific information provides several examples of selected articles with details related to assessment type, participant, country, and framework. Based on these criteria, this study tackles three main questions:

1. What is the general information about assessments utilized to explore STEM problem-solving?
2. What is the most common assessment employed in every discipline of STEM problem-solving?
3. What is the detailed information from selected articles related to participant, country, and framework applied in every discipline of STEM problem-solving?

Theoretical Background

Learning Domain in Problem-Solving

Assessment of problem-solving in the learning process is divided into several domains: cognitive, affective, and psychomotor. The cognitive domain emphasizes content knowledge or process skill predominantly related to mental (thinking) performance (Gao et al. 2020). Cognitive domain assessment is measured in summative/diagnostic and formative assessment. Bloom described the cognitive domain in levels from lower to higher, that is, remembering, understanding, applying, analyzing, evaluating, and creating. Three higher levels involve metacognition, which involves planning, monitoring, and evaluating problem-solving (Aurah, Cassady, and McConnell 2014).

The affective domain includes attitude, emotion, feelings, and behaviors (Gao et al. 2020). In this study, the affective domain refers to affective factors that impact STEM problem-solving. In addition, we will not include the psychomotor domain because problem-solving is a cognitive thinking process that applies knowledge and experience to come up with a solution (Annaggar and Tiemann 2020; Lertyosbordin et al. 2019; OECD 2013). Hence, cognitive aspects are emphasized in a problem-solving process.

Science, Technology, Engineering, and Mathematics Problem-Solving

Knowledge Integration in STEM Problem-Solving

The definition of STEM informs opinion about knowledge integration or combining two or more disciplines (Gao et al. 2020; Mücahit, Kurtuluş, and Bilen 2020). The disciplines in STEM are science, technology, engineering, mathematics, or subtopics of these areas (e.g., algebra in mathematics).

The four types of knowledge integration are monodisciplinary, multidisciplinary, interdisciplinary, and transdisciplinary (Gao et al. 2020). "Monodisciplinary" refers to one discipline or domain-specific (e.g., mathematics or sciences). Multidisciplinary refers to two or more disciplines (Samuels 2009). Interdisciplinary allows concepts, methods, or principles in at least two disciplines to be explicitly integrated (Gao et al. 2020). In contrast, transdisciplinary goes beyond disciplinary constraints in solving concrete problems or resolving project challenges (Gao et al. 2020; Samuels 2009). Therefore, the transdisciplinary approach focuses on a problem (or project) rather than each discipline. In the assessment area, the monodisciplinary method targets individual discipline assessment, whereas the interdisciplinary method assesses connections between disciplines, and the transdisciplinary method refers to assessment beyond disciplinary (Gao et al. 2020).

Framework in STEM Problem-Solving Assessment

The term "heuristic problem-solving" is a favored framework that involves four phases, namely, understanding a problem, devising a plan, carrying out the plan, and looking back. The framework in engineering is focusing on design-based concepts (Baele 2017; Jolly 2016; Kelley and Knowles 2016; Lin et al. 2020; Yata, Ohtani, and Isobe 2020). Engineering design connects science and mathematical concepts through designing prototypes and evaluation (Shahali et al. 2017; Shanta and Wells 2020). The assessment is typically done through project-based learning (PjBL) and problem-based learning (PBL) (Capraro and Corlu 2013). Another framework for engineering problem-solving, which involves a linear model and a cyclic model (e.g., Woods 2000), is closely related to heuristic problem-solving. These models are similar to scientific inquiry problem-solving in constructing a hypothesis, developing an empirical test, and drawing a conclusion.

Technology is the result when engineering involves the application of mathematics and science (Jolly 2016; Priemer et al. 2020). Technology can also be defined as physical tools or artifacts that ease human effort (Jolly 2016; Kelley and Knowles 2016).

Science in STEM problem-solving can be defined as an application of science concepts (Jolly 2016). However, it can also be described as a process of scientific inquiry, in which scientific reasoning is an integral skill (Kelley and Knowles 2016; Priemer et al. 2020; Yata, Ohtani, and Isobe 2020).

Mathematics in STEM education refers to the application of mathematics concepts, procedures, and facts described as mathematical thinking (Jolly 2016; Kelley and Knowles 2016). Researchers exploit the mathematical approach because it is closely related to scientific inquiry, such as inquiry-based mathematics education. The opposite of the inquiry concept in mathematical problem-solving is "proving," which uses a deductive approach to generate a conclusion that does not necessarily require empirical data (Priemer et al. 2020).

The first international assessment in problem-solving is PISA 2003 (i.e., creative problem-solving), which evolved to an individual computer-based assessment in 2012 that engaged more in authentic, real-life scenarios (e.g., complex problem-solving). There are four phases: exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting. Moreover, problem-solving is incorporated into mathematics and science literacy and renewed mathematical literacy with the need for twenty-first-century skills (OECD 2013).

The Trends in International Mathematics and Science Study (TIMSS) is also an international assessment that measures students' content knowledge in mathematics and science and thinking skills. The thinking skills are related to knowing, applying, and reasoning. These skills are important in problem-solving (Lindquits, Philpot, and Mulis 2019; Centurino and Jones 2017)

Methods

This study is an in-depth analysis based on a systematic review of key assessment resources. The study implemented a Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) to review STEM problem-solving assessments. The PRISMA process has four steps: (1) identifying articles according to general keywords; (2) screening abstract, title, and keywords according to specified criteria; (3) checking the eligibility of full articles; and (4) obtaining included criterion articles. Figure 1 shows a flowchart of the systematic review process of STEM problem-solving assessment.

The data resources identified from Google Scholar, ResearchGate, ProQuest, ScienceDirect, ERIC, DOAJ, Scopus, and JSTORE were published between January 2010 and December 2020. The articles included journal and other academic literature (conference paper, dissertation, and research data) published in English. It applies keywords, including problem-solving, assessment, science, mathematics, engineering, technology, test, and STEM listed in abstract, title, and keywords. The criteria for the selection of articles were as follows:

1. Focuses on (1) assessment of problem-solving in STEM, (2) availability of problem-solving framework, and (3) individual assessment.
2. Participant: (a) Middle school students (grades 6–10) and high school students (grades 11 and 12) from 11 to 19 years of age. Elementary school students are excluded from this study because STEM problem-solving are higher-order competencies suitable for measurement in the formal operational stage, (b) students without disabilities, and (c) no specifications for gender or ethnicity.
3. Empirical study (experimental study and nonexperimental study). The experimental study interventions are PBL, PjBL, and STEM problem-solving.

After screening activities, articles that meet these criteria were analyzed in detail. The articles included were analyzed according to an integration knowledge divided into four types: monodisciplinary, multidisciplinary, interdisciplinary, and transdisciplinary. Each discipline was categorized into cognitive (summative/diagnostic and formative), metacognitive, and affective. Descriptive statistics were applied to represent each category and describe the STEM assessment trend for each year. These articles were analyzed in detail by focusing on (1) instrument, (2) framework for problem-solving, (3) topic distribution, and (4) participant. To analyze article quality, indicators proposed by Protogerou and Hagger (2020), Downs and Black (1998), and Korstjen and Moser (2018) were applied for survey study, for experimental study, and for qualitative study, respectively.

The results of the data extraction, which apply the PRISMA method, are shown in Figure 1.

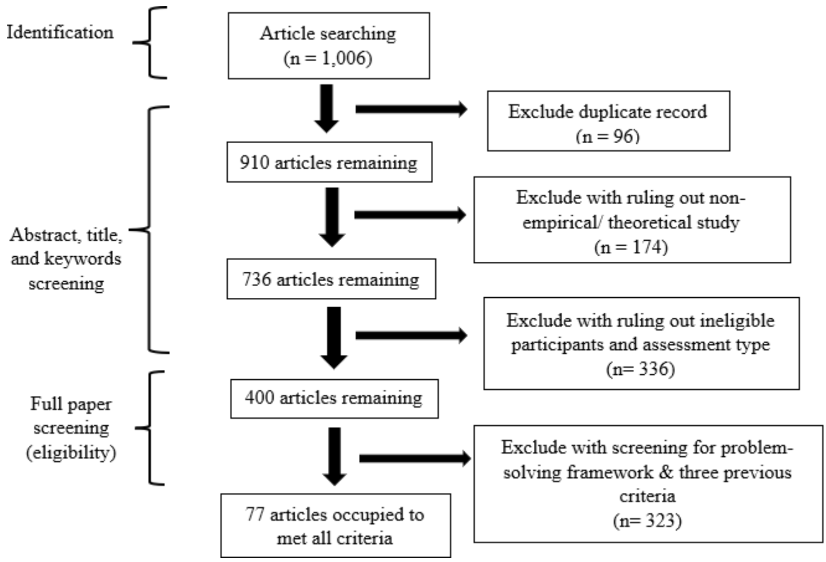


Figure 1: Result of Data Extraction

Results

Of the 1,006 articles found in the databases, seventy-seven satisfied the criteria. Most articles were excluded because of the eligibility of the participants and the problem-solving framework.

Figure 2 shows that the distribution of STEM problem-solving articles increases over time. Over the previous 5 years, 69 percent of all articles were published, with the highest distribution occurring in 2019-2020. These articles were focused on integrating more than one STEM discipline. It is clear that the assessment of problem-solving in STEM education remains a relevant topic for discussion.

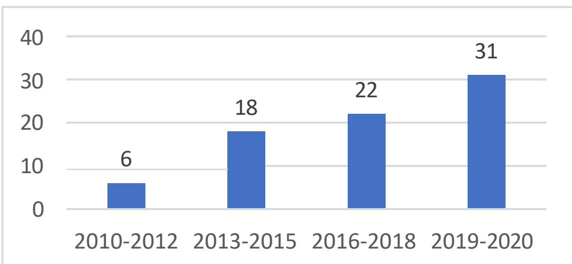


Figure 2: Distribution of Articles Over the Years

The highest distribution with respect to discipline (monodisciplinary [MD], interdisciplinary [ID], or transdisciplinary [TD]) was in mathematics, which was the subject of fifty-two of the seventy-seven articles. The second highest was science, with forty-six articles. Most articles were monodisciplinary in these two subjects (see Table 1 and Figure 3).

Table 1: Distribution of Articles Over Discipline and Domain of Learning

		MD	*ID*	*TD*
Cognitive	*Summative (Nonpractical)*	26	19	–
	Formative (Practical)	5	2	4
Metacognition		9	–	–
Affective		7	3	–
Cognitive Metacognition		1	–	–
Cognitive Affective		–	1	–

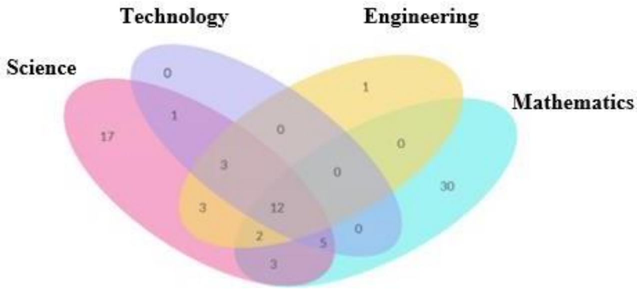


Figure 3: Distribution of Articles over the Discipline

Most articles were centered in the cognitive monodisciplinary area (n = 31). In terms of the distribution of participants, fifty-five studies were conducted with middle school students. Nineteen studies were concentrated mainly on high school students, and the studies for both level were very low in frequency (n = 3) (see Figure 4).

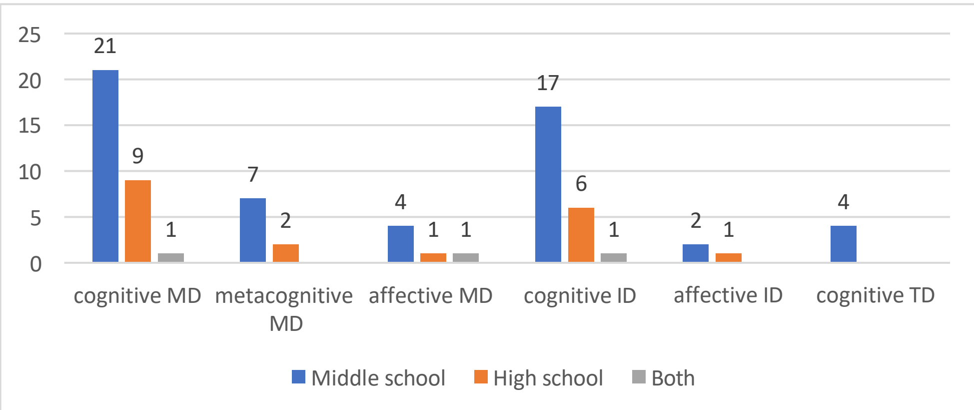


Figure 4: Distribution of Participants in Every Discipline

A specific finding refers to our research question "What are the common assessments utilized to explore problem-solving in each STEM discipline?" Figure 5 illustrates assessment types in every discipline in the cognitive, metacognitive, and affective areas. An article could apply more than one type of assessment.

The essay question (n = 25) is the most common type of assessment in cognitive MD. In cognitive ID, complex scenario problems were most prevalent (n = 11). Several studies were administered using computer-based assessment in cognitive problem-solving (Baele 2017; Chang 2010; Scherer and Tiemann 2012; Annaggar and Tiemann 2020; Eseryel, Ifenthaler, and Ge 2013; Eseryel et al. 2014; Scherer, Meßinger-Koppelt, and Tiemann 2014; Shute et al. 2016; Weise, Greiff, and Sparfeldt 2020). The questionnaire was the most common assessment in affective and metacognitive MD. However, metacognitive also used other in-depth assessments.

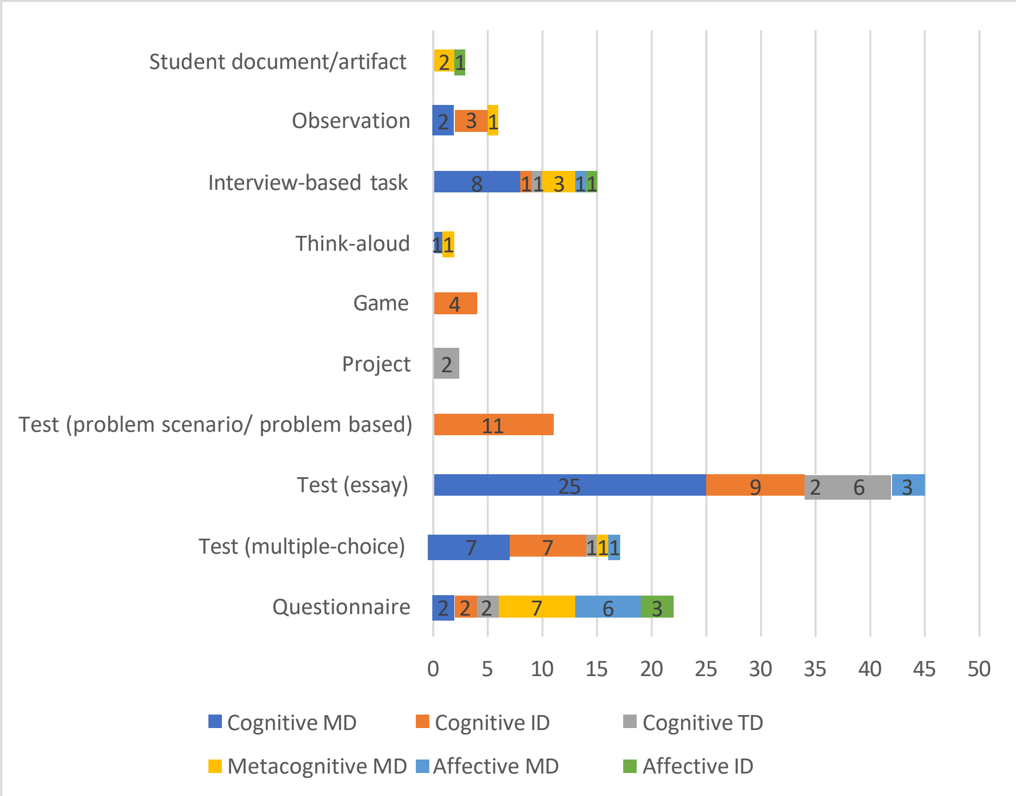


Figure 5: Assessment Types in Every Discipline and Domain of Learning

The distribution for the essay test has risen over the years, showing that this is still an appropriate type of assessment in cognitive STEM problem-solving. Furthermore, almost every year there is a study on the complex problem scenario test. The game and project assessments began to be applied in 2013 and 2015, respectively. In the metacognitive area, the “think aloud” assessment was applied in 2017.

In terms of research type, quantitative study ($n = 56$) was the most common type of study as compared with qualitative ($n = 13$) and mixed methods ($n = 8$). There was more nonexperimental research in the quantitative study ($n = 35$) than experimental research ($n = 21$). In the experimental study, various statistical tests were used (e.g., ANCOVA, MANOVA, t-test, ANOVA, N-gain, and descriptive statistics). By contrast, nonexperimental research mostly applied descriptive statistics and correlation tests (e.g., chi-squared, Pearson product moment, linear regression, ANOVA, and t-test). The psychometric analysis was also implemented in many studies to measure an instrument’s validity and reliability (e.g., Aiken’s V, Pearson’s product moment, expert validity, CFA, EFA, classical IRT, Rasch, and Cronbach alpha). Even so, ten studies did not report the validity or the reliability of their instruments. In selecting participants, thirty out of fifty-six studies in quantitative and mixed method employed non-randomized sampling.

In terms of article quality, risk of bias was measured, with thirty-seven articles deemed to be of high quality (low risk of bias). The survey studies mostly had general data to ensure that participants provided consent and were of low quality in terms of participant (sampling method, sample size, inclusion criteria, etc.). The qualitative studies mostly had vague findings and low quality in the reflexibility aspect, whereas the experimental studies were problematic with respect to internal validity. A more detailed example of the STEM problem-solving assessment will be described based on the type of assessment tools, topics, participants, data collection place, and framework.

Assessment of STEM Problem-Solving in the Cognitive MD Area

Researchers still utilized multiple-choice assessments to measure problem-solving in this area (e.g., Mustafa et al. 2017). A heuristic framework of problem-solving, PISA creative problem–solving framework, or scientific inquiry was used to measure students' problem-solving skills. One instrument, developed by Loh, Subramaniam, and Tan (2014) for tenth-grade students in Singapore on electrochemical cells, applied multiple choice and laid stress on student reasoning. It suggested that the instrument could be an alternative for measuring conceptual understanding. Although many studies focused on essay tests in cognitive monodisciplinary problem-solving, most emphasized closed-ended and routine problems. Assessment instruments were constructed by implementing a heuristic problem-solving framework (e.g., Du Toit and Du Toit 2013; İncebacak and Ersoy 2016; Karatas and Baki 2017), PISA problem-solving framework (e.g., Mustafa et al. 2017), or the Minnesota Assessment of Problem Solving (MAPS) (Docktor et al. 2016). The MAPS framework includes some requirements that focus on science concepts and mathematics procedures.

An example of a study conducted on essay-type closed-ended tests is that by İncebacak and Ersoy (2016), in simple geometry problems, with 15 percent finding it difficult to understand and evaluating the phases. Several studies utilized an open-ended essay test with more real-world problem-solving content, for example, a study by Karatas and Baki (2017) for Turkish seventh-grade students.

Practical problem-solving requires a more complex assessment. Such studies proposed assessment in engineering design or scientific inquiry. For example, Baele (2017) employed multiple-choice pretests and posttests to measure the content knowledge of hundred middle school students in engineering problem-solving. Moreover, Baele used open-ended tasks related to design-based problem-solving and conducted semistructured interviews related to students' understanding, difficulty, and problem-solving processes. The results showed that students still have difficulties in the phases related to recalling information and generating ideas.

Three studies used a few essay test problems ($n < 4$) (e.g., Du Toit and Du Toit 2013). The topics in the science area were related to physics (e.g., Docktor et al. 2015) and chemistry (e.g., Hidayat, Susilaningsih, and Kurniawan 2018; Loh, Subramaniam, and Tan 2014). The topics in mathematics were mostly related to number theory (e.g., İncebacak and Ersoy 2016; Karatas and Baki 2017), with other essays focusing on geometry, literacy, and algebra.

Several studies ($n = 4$) had a small sample size. Six studies with a few participants were concerned with qualitative research, mixed method, or experimental study, requiring in-depth investigation. Furthermore, one study only recorded the number of schools without giving an exact number of participants (e.g., Hidayat, Susilaningsih, and Kurniawan 2018). These studies were conducted in various countries (e.g., Malaysia, Brunei, Turkey, New York, Thailand, Italy, etc.), but most were carried out in Indonesia ($n = 13$).

Assessment of STEM Problem-Solving in the Metacognitive MD Area

The metacognitive aspect is emphasized in students' planning, monitoring, and reflecting. Although most of the studies applied in-depth assessment, several of them also used questionnaires and other assessment tools. All essay tests used in this area are closed-ended questions. Five out of six studies employed a Likert scale questionnaire (e.g., Özcan 2016; Velasquez and Bueno 2019), whereas the rest utilized a dichotomous scale.

An interesting study was conducted by Aurah, Cassady, and McConnell (2014), who implemented a metacognitive prompt dichotomous questionnaire embedded in a biology question for 2,138 students in the twelfth grade in Kenya that had a positive impact on problem-solving skills. The questionnaire focused on strategy, comprehension, reflection, and connection skills. Divinagracia (2018) implemented interviews, mathematics tests, and actual and video observations with eleven students (aged 15–17). The results showed that metacognitive skills emerged when the students solved mathematical problems.

The topics in metacognitive MD are related to number geometry and biology. Two studies had only a few participants because they called for a deep investigation (e.g., Divinagracia 2018). Furthermore, these studies were conducted in various countries (e.g., Indonesia, the Philippines, Turkey, and Kenya), but most were done in Turkey (n = 3).

Assessment of STEM Problem-Solving in the Affective MD and ID Areas

The studies in the affective MD area were carried out in several countries (e.g., Brunei, the United States, Estonia, Turkey), but most of them were organized in Turkey (n = 4). Six out of seven studies applied a Likert scale questionnaire to examine any link between affective and problem-solving. Several affective factors were correlated with students' problem-solving skills, such as anxiety, attitude, self-efficacy, self-perception, and interest in problem-solving (e.g., Guven and Cabakcor 2013; Soobard and Rannikmäe 2014).

A study by Guven and Cabakor (2013) explored affective factors and academic achievement influencing mathematical problems by using a heuristic mathematics problem-solving test and questionnaires. It found a high correlation between academic achievement and mathematical problem-solving but only a moderate correlation between attitude toward mathematics, self-efficacy, and anxiety. Soobard and Rannikmäe (2014) investigated a large Estonian sample of tenth-grade students' perceptions about their competence in scientific process skills and science future careers. The results showed that students' self-perception was higher in recognizing problems, drawing conclusions, using figures and tables as source information, and making evidence-based decisions. Moreover, students who had more negative self-perceptions about science tend to be less interested in a science career or learning scientific problem-solving.

The main assessment tool for the interdisciplinary affective category was a questionnaire. Two out of three studies focused on students interested in STEM (Mücahit, Kurtuluş, and Bilen 2020; Shahali et al. 2019). One large-scale study by Mücahit, Kurtuluş, and Bilen (2020) implemented a five-scale Likert questionnaire about STEM attitude and reflective thinking in problem-solving for fifth- and eighth-grade students in Turkey. The findings revealed (1) a strong positive correlation between STEM attitude and reflective skills in problem-solving and (2) reflective skills in problem-solving predicted STEM attitude.

Assessment of STEM Problem-Solving in the Cognitive ID and TD Areas

Essay tests in the cognitive ID area were of both open-ended (e.g., Chusinkunawut et al. 2018) and closed-ended types (e.g., Tasir et al. 2018), with an ill-defined problem. A variation of a heuristic problem-solving framework was used. A study of forty-one tenth-grade students in Thailand by Chusinkunawut et al. (2018) applied an essay test with a rubric and test-related design of problem-solving. This assessment focused on an open-ended problem (with several given constraints) about electrochemistry. The results suggested that a rubric and a test are suitable for gauging students' mental models in applying design-based problem-solving. However, 40 percent of students still could not apply the scientific concepts they had learned to real-world problems.

Further, seven out of eleven studies administered complex scenario and game assessments using computer-based conclusions. The framework used was engineering design or the problem-solving framework of MAPS. Annaggar and Tiemann (2020) developed a computer-based game assessment called ALCHEMIST for a study of seventy-five eighth-grade students in Germany, highlighting STEM to support chemistry problem-solving. They emphasized problem-solving skills such as understanding and characterization, representation, solving problems, and reflection. The ChemLabBuilder assessment applies a framework similar to ALCHEMIST with a procedure similar to MicroDYN, but for high school students. Another study, by Shute et al. (2016), constructed a stealth assessment using a game called "Use Your Brainz" for seventh-grade American students. The validity of these assessment tools had a moderate correlation with MicroDYN. Another assessment in solar system topics was used by researchers for postsecondary students in the United States and related to designing a living habitat of an Earth-like planet for vehicle production and precipitation gauge (Chang 2010; Eseryel et al. 2014; Schwortz and Burrows 2020).

Shanta (2019) and Shanta and Wells (2020) developed a paper-based evaluation based on MAPS rubric and engineering-based design called "design-no-make challenge." The assessment allows the utilization of design skills related to a suitable size and price pump to supply water within specified constraints. It could be used to test American high school preengineering students. Shanta and Wells (2020) conducted a further study on twelfth-grade students and found that students' skills in applying physics and mathematics correlated with overall success in problem-solving.

Assessment in the cognitive ID area addressed topics closely related to everyday phenomena. Most science topics fall under electricity, energy, environment (plant, water, chemical substance, etc.), movement, and the solar system (e.g., Chang 2010; Eseryel et al. 2014; Eseryel, Ifenthaler, and Ge 2013; Shanta 2019; Shanta and Wells 2020). Mathematics commonly uses ratio, number, data, and statistics (e.g., Annaggar and Tiemann 2020; Shanta 2019; Shanta and Wells 2020; Shute et al. 2016). With respect to sample size, ten studies had only a few participants (less than seventy). However, four studies were experimental (e.g., Lertyosbordin et al. 2019; Shanta and Wells 2020), and the remaining six were pilot studies (e.g., Chusinkunawut et al. 2018; Shute et al. 2016). The studies were administered primarily in the United States ($n = 7$), whereas others were conducted in Germany, Taiwan, Thailand, Vietnam, and Indonesia.

Studies in the cognitive TD area are related to robotics (e.g., Robinson 2016) and projects (e.g., Lin et al. 2020; Shahali et al. 2017) that emphasized an engineering-based design framework. A study by Robinson (2016) used a robotics and engineering design curriculum to test the STEM problem-solving capacity of fifty-four American eighth-grade students. A different study looked at project-based assessment in science education through solar-powered vehicle design (Shahali et al. 2017) and egg protection devices (Lin et al. 2020). The studies were conducted on Taiwanese and Malaysian middle school students and assessed through a complex problem test.

A more complex assessment, promoted by Yu, Fan, and Lin (2015), was carried out on 103 eighth-grade Taiwanese students. They designed a complex problem-solving assessment through a three-stage exercise: (1) detective films, allowing students to deduce knowledge, tools, and scientific principles after watching the video; (2) context simulation stage, giving students an opportunity to solve a problem related to the film; and (3) project design stage, enabling students to design a capable vehicle based on the problem on the context simulation stage. In another challenge, students were required to solve problems by creating a plan to rescue people in a film context.

Discussion and Implications

Assessment of problem-solving is a fundamental aspect of STEM education, as shown by the increase in studies over the past five years. This review shows that STEM problem-solving assessments focus mostly on cognitive areas, both monodisciplinary and interdisciplinary. No results were found in the multidisciplinary area because assessment in STEM underscores integration of two or more disciplines (Mücahit, Kurtuluş, and Bilen 2020; Shahali et al. 2017). Furthermore, science and mathematics are still key disciplines in these studies, but several studies in the interdisciplinary and transdisciplinary areas have attempted to integrate them using engineering and technology.

Various types of assessment in STEM problem-solving have been described. However, each type of assessment has its advantages and drawbacks, which are presented in Table 2.

Table 2: The Advantages and Drawbacks of Assessment in Cognitive STEM Problem-Solving

Type of Assessment	*Advantages*	*Drawbacks*
Questionnaire	Trains students' metacognition to assess their knowledge about problem-solving	▪ Assesses their perception regarding their skills in problem-solving ▪ In the interdisciplinary and transdisciplinary problem-solving articles, they mostly fail to integrated disciplines into the questionnaire items.
Multiple-Choice Test	▪ Easy to assess ▪ The scoring method is objective ▪ In the interdisciplinary and transdisciplinary problem-solving studies, technology is used to make the tasks more interactive.	▪ Assesses the product of thinking rather than the process itself ▪ In the monodisciplinary, the type of tasks is static.
Essay Test	▪ Measures students' thinking process. Mostly, the tasks include all the problem-solving steps. ▪ Gives students more space to explore problem-solving skills ▪ In the interdisciplinary and transdisciplinary studies, the tasks are ill-defined and explore higher-order thinking.	▪ In monodisciplinary, the type of task is static, not authentic, mostly closed-ended, and simple constraint. ▪ The reliability evidence of the tools mostly didn't report.
*Essay Test (Complex Problem Scenario)**	▪ More authentic because it is related to the real issues (e.g., environmental issues) ▪ Makes students' thinking process visible ▪ More interactive with the help of technology ▪ The reliability of the tasks is reported acceptable	▪ The interdependency of the task with reading skills
*Game**	Creates a pleasure environment	The flexibility of the use is questionable
*Project***	▪ Proposes learning by doing ▪ The tasks are authentic problems	▪ Needs high level of prior knowledge ▪ Challenging in assessing. The scoring rubric is not clearly described. ▪ Research oriented
Think Aloud	Measures thinking process accurately	The method is interdependent with other methods. In the selected study, the task used in the think aloud is routine and far from being a problem-solving concept.
Interview-Based Task	Confirms students' thinking process and understanding	Difficult to interpret the result because some reviewed studies failed to describe in in-depth analysis
Observation	Measures problem-solving skills in natural setting	Challenging in creating scoring rubric

*Only appears in cognitive interdisciplinary; **Only appears in cognitive transdisciplinary
Source: Amalina and Vidákovich

The cognitive MD assessment area is dominated by closed-ended essay tests but its reliability is questionable. Researchers mostly employed a heuristic problem-solving framework or other related frameworks. This type of assessment contrasts with the premise that everyday problems are typically open ended, ill-defined, and complex (Chusinkunawut et al. 2018). The heuristic problem-solving framework is suitable for assessment in the monodisciplinary area. Although it only measures a discipline, it is a core assessment of students' domain-specific problem-solving ability (Chang 2010). Despite its drawback, the essay test is recommended for monodisciplinary problem-solving because it can explore the process of students' thinking. It is suggested that the essay test be combined with think aloud or interview-based tasks. The authentic, interactive, open-ended and nonroutine essay tests are recommended. The static type of questions should be avoided.

Questionnaire and observation are not suggested for use in measuring cognitive STEM problem-solving because the thinking process cannot be observed directly and doesn't appear through behavior (Lertyosbordin et al. 2019). In the cognitive MD area, multiple-choice can be an assessment option because it only measures a discipline. However, an interdependent multiple-choice item that represents a process is suggested. In a practical/formative monodisciplinary assessment, observation was conducted mostly with a framework of engineering design and scientific inquiry.

The essay scenario test is the most commonly used form of assessment in cognitive ID STEM problem-solving. In addition, the high psychometric evidence is reported (Annaggar and Tiemann 2020; Shute et al. 2016; Shanta 2019). However, some of the tasks need student reading comprehension. Splitting the task by giving metacognition prompting questions can be the solution (Shanta 2019). The use of games as an assessment tool in the reviewed studies was a part of the essay scenario test. Hence, the application of both essay scenario task and games as a comprehensive form of assessment in the interdisciplinary is suggested (Annaggar and Tiemann 2020; Eseryel, Ifenthaler, and Ge 2013; Eseryel et al. 2014; Shute et al. 2016).

Although STEM problem-solving is beginning to focus more on both formative and summative assessment, it is still possible to do a paper-based evaluation in a nonpractical cognitive area as a summative assessment (e.g., Chusinkunawut et al. 2018). In addition, the view of interdisciplinary STEM problem-solving assessment not only focuses on science and mathematics but also includes engineering and technology (Kelley and Knowles 2016). This has led to several interdisciplinary studies with an engineering design and technology approach through complex design-based tests, games, and complex problem scenarios.

A transdisciplinary area of problem-solving assessment that focuses on individual assessment is rare. This is because the transdisciplinary assessment focuses mostly on group assessment. However, it is still possible to examine individual assessment (Capraro and Corlu 2013). The transdisciplinary area is mostly recommended to be evaluated using project-based assessment (Gao et al. 2020; Samuels 2009), and the assessment in this area is a complex task related to robotics and science projects.

Most studies applied number, quantity, and ratio in the mathematics area, whereas electricity, environment, energy, earth, and solar system are utilized for the science area. This comes from recommendations made by the OECD in PISA assessment (OECD 2013). Not only the topic but also the quantity of test items is considered important in developing an instrument. With insufficient questions, the consistency of students' answers presents analytical challenges. There are still studies that used only a few items (e.g., an item only) in an essay test.

Assessment type to measure affective and metacognitive skills in STEM problem-solving were reviewed. However, we display the advantages and drawbacks to help guide the future direction in this area (see Table 3).

Table 3: The Advantages and Drawbacks of Assessment Type in Affective and Metacognitive STEM Problem-Solving

Type of Assessment	*Advantages*	*Drawbacks*
Questionnaire	▪ Efficient regarding time and large scale used ▪ Fulfills the content validity based on their frameworks ▪ Strong psychometric evidence	Interdependency in participant perception or beliefs
*Multiple-Choice Test**	Scoring method is objective	Restricts student thinking
*Essay Test**	Students' response is not restricted	▪ The scoring rubric is questionable ▪ Time-consuming
*Artifact***	Assesses affective in a natural environment	Interdependency with other assessment method
Interview-Based Task	In-depth analysis of affective skills	Time-consuming and challenging in reporting the result
*Think Aloud****	Monitors student thinking that sometime is not visible in written test	Challenging in coding system
*Observation****	Measures in natural setting	Metacognitive skills are not visible in a natural setting

*Only appears in cognitive interdisciplinary; **Only appears in cognitive transdisciplinary;
***Only appears in metacognitive STEM problem-solving
Source: Amalina and Vidákovich

The most common and appropriate tool for measuring affective in STEM problem-solving is the questionnaire because this type of assessment is effective and efficient for a large student sample (Guven and Cabakcor 2013). However, in the affective ID area, the way researchers integrate among disciplines needs to be considered because interdisciplinary is not put together among disciplines independently (Gao et al. 2020). An interview can be an option together with a questionnaire to figure comprehensively students affective skills in STEM problem-solving. Although there was assessment in affective ID area, only a few articles met the criteria. This is in accord with the finding by Mücahit, Kurtuluş, and Bilen (2020) that studies in affective STEM interdisciplinary related to problem-solving are rarely found owing to integration difficulties.

In metacognitive MD STEM problem-solving, the most popular assessment is the questionnaire and other in-depth forms of assessment (e.g., interview, observation, or think aloud). However, it is necessary to utilize more than one assessment tool because measuring metacognitive knowledge is challenging and involves implicit knowledge and metamemory (Scherer and Tiemann 2012). It is well established that no study has done any assessment in the metacognitive interdisciplinary category because it includes cognitive factors.

Participants in the fifty-five studies were middle school students; this is the initial stage when students begin to think critically (Gao et al. 2020; Kelley and Knowles 2016). The studies were conducted mostly in Indonesia (MD cognitive area), Turkey (MD metacognitive and affective area), and the United States (ID cognitive area). Several studies had fewer than thirty participants owing to the type of studies (pilot study or qualitative study) that need in-depth analysis.

The following are challenges in developing and assessing STEM problem-solving based on the reviewed studies.

1. All reviewed studies mentioned their framework in developing and measuring STEM cognitive problem-solving. However, the frameworks in STEM ID cognitive problem-solving were not clearly articulated. There is no standardized framework. Researchers only mentioned that they utilized a combination of frameworks in each STEM discipline but didn't mention how they combined them. Because combining several frameworks is not an easy task, it is necessary to have a standardized framework.
2. Interdisciplinary in STEM problem-solving assessment is neither explicitly theorized nor well articulated. It is not merely a matter of taking several disciplines together but needs intentional and specific consideration of the connection across discipline based on framework and curriculum because only several specific topics in each of STEM discipline can be integrated.
3. Among the studies we reviewed, many reported assessments are research oriented because several types of assessment are time-consuming. Teachers don't have time to conduct these assessment tools in a classroom setting.
4. Assessment of STEM problem-solving can't rely only on a single tool given the complexity of the skills. The use of a constructed type of assessment is recommended, but the psychometric evidence for this type of assessment is mostly difficult to obtain.

This study can be improved. Because gray literature (conference papers and dissertation reports) and any type of journal (without concerning their impact factor) have been included in the current study, there is a limitation in the quality of articles. Furthermore, the reliability and validity of the reported articles are not described. Lastly, this study focuses only on an individual assessment, affecting the number of articles in transdisciplinary areas.

This study is important in that it helps teachers to select the best, standardized, and most widely used framework to develop a test that assesses student problem-solving skills in the STEM area based on the drawbacks and advantages listed by the authors. Hence, teachers can explore every step of students' problem-solving skills. The result can be a basis for improving students' problem-solving skills and deciding the next teaching and learning strategy. Another important suggestion for teachers, based on this study's findings, is that they can select the appropriate available assessment tools based on the assessments' psychometric results. However, teachers are recommended to revise or adapt the content of available assessments based on students' needs.

The results of this study have important implications for other researchers to investigate the trend of the assessment in STEM education, including the trend of assessment tools, participants, topics, and the countries where the studies were conducted. Hence, it can be used to strengthen their research claims or research background. Moreover, they can find gaps in these trends for their future research. Furthermore, they can use this study as a basis for deciding the best instrument or assessment tools to be applied in every discipline, or they can utilize this review study to select the assessment for development or adaptation. The selected limitation for this study will be useful for other researchers who propose to conduct a review study to fill the gap.

Conclusion and Further Direction

The assessment of STEM problem-solving in domain-specific categories was still more prevalent than in domain-general problem-solving, which focused mainly on science and mathematics. However, several studies changed the approach by embedding technology and engineering as an integrator by applying engineering design, scientific inquiry, and technology application. The PISA assessment contents were still widely used topics in STEM problem-solving targeted at middle school students.

Most of the assessments in the cognitive MD area focused on conducting essay tests with an applied heuristic framework or related framework. Compared with the monodisciplinary, the interdisciplinary area was more complex because it emphasized the constructed type of assessment that allows students to explore their metacognitive skills (complex problem scenario, game). Furthermore, assessment in transdisciplinary problem-solving focused on projects. The assessment of metacognitive monodisciplinary was dealt with through an in-depth assessment. Assessments of affective factors still used questionnaires, but in interdisciplinary assessment, it is required for further study in the development of assessment tools.

Because of the trend of setting STEM problem-solving assessment within the monodisciplinary view, it is entirely possible to develop a tool for interdisciplinary problem-solving. It will be focused on the complex scenario problem in the contexts recommended by the OECD. Moreover, affective traits will comprise factors in problem-solving that need to be encouraged.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This research received no external funding.

REFERENCES

Annaggar, Amany, and Rudiger Tiemann. 2020. “A Video Game as an Assessment Tool of Problem-Solving Competence.” In *Gesellschaft Fur Didactik Der Chemie Und Physik (GDCP)* [The Society for Didactics of Chemistry and Physics], edited by Sebastian Habig, 377–380. Essen, DE: Universität Duisburg-Essen.

Aurah, Catherine M., Jerrell C. Cassady, and Tom J. McConnell. 2014. “Predicting Problem Solving Ability from Metacognition and Self-Efficacy Beliefs on a Cross Validated Sample.” *British Journal of Education* 2 (1): 49–72. https://www.eajournals.org/journals/british-journal-of-education-bje/vol-2-issue-1march-2013/predicting-problem-solving-ability-metacognition-self-efficacy-beliefs-cross-validated-sample-2/.

Baele, Loren C. 2017. “Middle School Engineering Problem Solving Using Traditional vs. E-PBL Module Instruction.” EdD diss., Aurora University.

Capraro, Robert M., and Sencer C. Corlu. 2013. "Changing Views on Assessment for STEM Project-Based Learning." In *STEM Project-Based Learning*, edited by Robert M. Capraro, M. M. Capraro, and J. R. Morgan, 109–118. Rotterdam, NL: Sense Publishers.

Centurino, Victoria, and Lee Jones. 2017. "TIMSS 2019 Science Framework." *TIMSS 2019 Assessment Frameworks*, edited by I. V. S. Mullis and M. O. Martin, 29–55. Boston: Boston College.

Chang, Chun-Yen. 2010. "Does Problem Solving = Prior Knowledge + Reasoning Skills in Earth Science? An Exploratory Study." *Research in Science Education* 40 (2): 103–116. https://doi.org/10.1007/s11165-008-9102-0.

Chusinkunawut, Krislada, Kulthida Nugultham, Tussatrin Wannagatesiri, and Witat. Fakcharoenphol. 2018. "Problem Solving Ability Assessment Based on Design for Secondary School Students." *International Journal of Innovation in Science and Mathematics Education* 26 (3): 1–20. https://openjournals.library.sydney.edu.au/CAL/article/view/12413.

Divinagracia, Lovella S. 2018. "Metacognitive Processes in Successful Mathematical Problem Solving." *Future Ed Special Edition: Multidisciplinary Research Journal* 5 (1): 70–86. https://www.dlsl.edu.ph/wp-content/uploads/2020/04/7.pdf.

Docktor, Jennifer L., Jay Dornfeld, Evan Frodermann, Kenneth Heller, Leonardo Hsu, Koblar Alan Jackson, Andrew Mason, Qing X. Ryan, and Jie Yang. 2016. "Assessing Student Written Problem Solutions: A Problem-Solving Rubric with Application to Introductory Physics." *Physical Review Physics Education Research* 12 (1): 1–18. https://doi.org/10.1103/PhysRevPhysEducRes.12.010130.

Docktor, Jennifer L., Natalie E. Strand, Jose P. Mestre, and Brian H. Ross. 2015. "Conceptual Problem Solving in High School Physics." *Physical Review Special Topics – Physics Education Research* 11 (2): 1–13. https://doi.org/10.1103/PhysRevSTPER.11.020106.

Downs, Sara H., and Nick Black. 1998. "Downs and Black 1998 the Feasibility of Creating a Checklist for the Assessment of the Methodological Quality Both of Randomised and Non-Randomised Studies of Health.Pdf." *Journal Epidemiol Community Health* 52:377–384. https://doi.org/10.1136/jech.52.6.377.

Du Toit, S., and G. Du Toit. 2013. "Learner Metacognition and Mathematics Achievement during Problem-Solving in a Mathematics Classroom." *Journal for Transdisciplinary Research in Southern Africa* 9 (3): 505–518. https://doi.org/10.4102/td.v9i3.194.

Eseryel, Deniz, Dirk Ifenthaler, and Xun Ge. 2013. "Validation Study of a Method for Assessing Complex Ill-Structured Problem Solving by Using Causal Representations." *Educational Technology Research and Development* 61 (3): 443–463. https://doi.org/10.1007/s11423-013-9297-2.

Eseryel, Deniz, Victor Law, Dirk Ifenthaler, Xun Ge, and Raymond Miller. 2014. "An Investigation of the Interrelationships between Motivation, Engagement, and Complex Problem Solving in Game-Based Learning." *Educational Technology and Society* 17 (1): 42–53. https://www.jstor.org/stable/jeductechsoci.17.1.42.

Gao, Xiaoyi, Peishan Li, Ji Shen, and Huifang Sun. 2020. "Reviewing Assessment of Student Learning in Interdisciplinary STEM Education." *International Journal of STEM Education* 7 (1): 1–14. https://doi.org/10.1186/s40594-020-00225-4.

Greiff, Samuel, and Joachim Funke. 2009. "Measuring Complex Problem Solving: The MicroDYN Approach." *Main*, September: 1–6. http://archiv.ub.uni-heidelberg.de/volltextserver/15502/.

Guven, Bulent, and Buket Ozum Cabakcor. 2013. "Factors Influencing Mathematical Problem-Solving Achievement of Seventh Grade Turkish Students." *Learning and Individual Differences* 23 (1): 131–137. https://doi.org/10.1016/j.lindif.2012.10.003.

Hidayat, Taufiq, Endang Susilaningsih, and Cepi Kurniawan. 2018. "The Effectiveness of Enrichment Test Instruments Design to Measure Students' Creative Thinking Skills and Problem-Solving." *Thinking Skills and Creativity* 29 (February): 161–169. https://doi.org/10.1016/j.tsc.2018.02.011.

İncebacak, Belgin Bal, and Esen Ersoy. 2016. "Problem Solving Skills of Secondary School Students." *China-USA Business Review* 15 (6): 275–285. https://doi.org/10.17265/1537-1514/2016.06.002.

Jolly, Anne. 2016. *STEM by Design: Strategies and Activities for Grades 4-8*. New York: Routledge.

Karatas, Ilhan, and Adnan Baki. 2017. "The Effect of Learning Environments Based on Problem Solving on Students' Achievements of Problem Solving." *International Electronic Journal of Elementary Education* 5 (3): 249–267. https://www.iejee.com/index.php/IEJEE/article/view/25.

Kelley, Todd R., and J. Geoff Knowles. 2016. "A Conceptual Framework for Integrated STEM Education." *International Journal of STEM Education* 3 (1): 11. https://doi.org/10.1186/s40594-016-0046-z.

Korstjens, Irene, and Albine Moser. 2018. "Series: Practical Guidance to Qualitative Research. Part 4: Trustworthiness and Publishing." *European Journal of General Practice* 24 (1): 120–124. https://doi.org/10.1080/13814788.2017.1375092.

Lertyosbordin, Chacharin, Sorakrich Maneewan, Sakesun Yampinij, and Kutinda Thamwipat. 2019. "Scoring Rubric of Problem-Solving on Computing Science Learning." *International Education Studies* 12 (8): 26. https://doi.org/10.5539/ies.v12n8p26.

Lin, Kuen-Yi, Hsien-Sheng Hsiao, P. John Williams, and Yu-Han Chen. 2020. "Effects of 6E-Oriented STEM Practical Activities in Cultivating Middle School Students' Attitudes toward Technology and Technological Inquiry Ability." *Research in Science and Technological Education* 38 (1): 1–18. https://doi.org/10.1080/02635143.2018.1561432.

Lindquits, Mary, Ray Philpot, and Ina Mulis. 2019. "TIMSS 2019 Mathematics Assessment Framework." *TIMSS 2019 Assessment Frameworks*, edited by I. V. S. Mullis and M. O. Martin. 13–24. Boston: Boston College.

Loh, Andrian Sin Loy, R. Subramaniam, and Kim Chwee Daniel Tan. 2014. "Exploring Students' Understanding of Electrochemical Cells Using an Enhanced Two-Tier Diagnostic Instrument." *Research in Science and Technological Education* 32 (3): 229–250. https://doi.org/10.1080/02635143.2014.916669.

Mücahit, Köse, Muhammed Akif Kurtuluş, and Kadir Bilen. 2020. "The Relationship of STEM Attitudes and Reflective Thinking Skills on Problem-Solving." *OPUS International Journal of Society Researches* 16 (27): 76–93. https://doi.org/10.26466/opus.667004.

Mustafa, Norazla, Mohd Nihra Haruzuan Mohammad Said, Zaleha Ismail, and Zaidatun Tasir. 2017. "Mathematical Problem Solving Skills among High Achiever Students." *Advanced Science Letters* 23 (8): 7494–7498. https://doi.org/10.1166/asl.2017.9506.

OECD. 2013. *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy.* OECD Publishing.

Özcan, Zeynep Cigdem. 2016. "The Relationship between Mathematical Problem-Solving Skills and Self-Regulated Learning through Homework Behaviours, Motivation, and Metacognition." *International Journal of Mathematical Education in Science and Technology* 47 (3): 408–420. https://doi.org/10.1080/0020739X.2015.1080313.

Priemer, Burkhard, Katja Eilerts, Andreas Filler, Niels Pinkwart, Bettina Rösken-Winter, Rudiger Tiemann, and Annette Upmeier Zu Belzen. 2020. "A Framework to Foster Problem-Solving in STEM and Computing Education." *Research in Science and Technological Education* 38 (1): 105–130. https://doi.org/10.1080/02635143.2019.1600490.

Protogerou, Cleo, and Martin S. Hagger. 2020. "A Checklist to Assess the Quality of Survey Studies in Psychology." *Methods in Psychology* 3 (June): 100031. https://doi.org/10.1016/j.metip.2020.100031.

Robinson, Norman. 2016. “A Case Study Exploring the Effects of Using an Integrative STEM Curriculum on Eighth Grade Students’ Performance and Engagement in the Mathematics Classroom.” PhD diss., Georgia State University.

Samuels, Boba M. 2009. “Can the Differences between Education and Neuroscience Be Overcome by Mind, Brain, and Education?” *Mind, Brain, and Education* 3 (1): 45–55. https://doi.org/10.1111/j.1751-228X.2008.01052.x.

Scherer, Ronny, Jenny Meßinger-Koppelt, and Rudiger Tiemann. 2014. “Developing a Computer-Based Assessment of Complex Problem Solving in Chemistry.” *International Journal of STEM Education* 1 (1): 1–15. https://doi.org/10.1186/2196-7822-1-2.

Scherer, Ronny, and Rudiger Tiemann. 2012. “Factors of Problem-Solving Competency in a Virtual Chemistry Environment: The Role of Metacognitive Knowledge about Strategies.” *Computers and Education* 59 (4): 1199–1214. https://doi.org/10.1016/j.compedu.2012.05.020.

Schwortz, Andria C., and Andrea C. Burrows. 2020. “Authentic Science Experiences with STEM Datasets: Post-Secondary Results and Potential Gender Influences.” *Research in Science and Technological Education* 39 (3): 347–367. https://doi.org/10.1080/02635143.2020.1761783.

Shahali, Edy Hafizan Mohd, Lilia Halim, Mohamad Sattar Rasul, Kamisah Osman, and Nurazidawati Mohamad Arsad. 2019. “Students’ Interest Towards STEM: A Longitudinal Study.” *Research in Science and Technological Education* 37 (1): 71–89. https://doi.org/10.1080/02635143.2018.1489789.

Shahali, Edy Hafizan Mohd, Lilia Halim, Mohamad Sattar Rasul, Kamisah Osman, and Mohd Afendi Zulkifeli. 2017. “STEM Learning through Engineering Design: Impact on Middle Secondary Students’ Interest Towards STEM.” *Eurasia Journal of Mathematics, Science and Technology Education* 13 (5): 1189–1211. https://doi.org/10.12973/eurasia.2017.00667a.

Shanta, Susheela. 2019. “Developing and Assessing Authentic Problem-Solving Skills in High School Pre-Engineering Students.” ASEE Annual Conference and Exposition, Conference Proceedings, Tampa, FL, 2019. https://doi.org/10.18260/1-2--32637.

Shanta, Susheela, and John G. Wells. 2020. “T/E Design Based Learning: Assessing Student Critical Thinking and Problem Solving Abilities.” *International Journal of Technology and Design Education* 32:267–285. https://doi.org/10.1007/s10798-020-09608-8.

Shute, Valerie J., Lubin Wang, Samuel Greiff, Weinan Zhao, and Gregory Moore. 2016. “Measuring Problem Solving Skills via Stealth Assessment in an Engaging Video Game.” *Computers in Human Behavior* 63:106–117. https://doi.org/10.1016/j.chb.2016.05.047.

Soobard, Regina, and Miia Rannikmäe. 2014. “Upper Secondary Students’ Self-Perceptions of Both Their Competence in Problem Solving, Decision Making and Reasoning within Science Subjects and Their Future Careers.” *Journal of Baltic Science Education* 13 (4): 544–558. https://doi.org/10.33225/jbse/14.13.544.

Tasir, Zaidatun, Mohd Nihra Mohamed Mohamad, Zaleha Ismail, and Norazla Mustafa. 2018. “Development and Validation of Problem Solving Task Based-Integrated STEM.” *Proceedings—2018 6th International Conference on Learning and Teaching in Computing and Engineering, LaTiCE 2018*, 17–23. https://doi.org/10.1109/LaTICE.2018.00011.

Velasquez, Quezco F., and David Cababaro Bueno. 2019. “Metacognitive Skills in Problem Solving of Senior High School STEM Strand Students.” *Institutional Multidisciplinary Research and Development* 2 (June): 124–129. https://www.researchgate.net/profile/David_Cababaro_Bueno/publication/334316016_Metacognitive_Skills_in_Problem_Solving_of_Senior_High_School_STEM_Strand_Students/links/5d23e97ba6fdcc2462ce00f6/Metacognitive-Skills-in-Problem-Solving-of-Senior-High-School-STEM-Strand-Students.pdf.

Weise, Julius J., Samuel Greiff, and Jorn R. Sparfeldt. 2020. "The Moderating Effect of Prior Knowledge on the Relationship between Intelligence and Complex Problem Solving—Testing the Elshout-Raaheim Hypothesis." *Intelligence* 83 (October): 101502. https://doi.org/10.1016/j.intell.2020.101502.

Woods, Donald R. 2000. "An Evidence-Based Strategy for Problem Solving." *Journal of Engineering Education* 89 (4): 443–459. https://doi.org/10.1002/j.2168-9830.2000.tb00551.x.

Yata, Chikahiko, Tadashi Ohtani, and Masataka Isobe. 2020. "Conceptual Framework of STEM Based on Japanese Subject Principles." *International Journal of STEM Education* 7 (1): 12. https://doi.org/10.1186/s40594-020-00205-8.

Yu, Kuang Chao, Szu Chun Fan, and Kuen Yi Lin. 2015. "Enhancing Students' Problem-Solving Skills through Context-Based Learning." *International Journal of Science and Mathematics Education* 13 (6): 1377–1401. https://doi.org/10.1007/s10763-014-9567-4.

ABOUT THE AUTHORS

Ijtihadi Kamilia Amalina, MEd: PhD Student, Doctoral School of Education, University of Szeged, Szeged, Hungary

Professor Tibor Vidákovich: Professor, Institute of Education, University of Szeged, Szeged, Hungary

The International Journal of Assessment and Evaluation is one of ten thematically focused journals in the collection of journals that support The Learner Research Network—its journals, book series, conference, and online community. The journal investigates the dimensions of educational measurement.

As well as articles of a traditional scholarly type, this journal invites presentations of practice—including documentation of measurement practices and exegeses of the effects of those practices.

The International Journal of Assessment and Evaluation is a peer-reviewed, scholarly journal.

ISSN: 2327-7920

Study 2: An integrated STEM-based mathematical problem-solving test: developing and reporting psychometric evidence.

(Amalina & Vidákovich, 2022a)

Journal on Mathematics Education
Volume 13, No. 4, 2022, pp. 587-604

An integrated STEM-based mathematical problem-solving test: Developing and reporting psychometric evidence

Ijtihadi Kamilia Amalina[1], Tibor Vidákovich[2]

[1]Doctoral School of Education, University of Szeged, Hungary
[2]Institute of Education, University of Szeged, Hungary
*Correspondence: ijtihadi.kamilia.amalina@edu.u-szeged.hu

Received: 15 October 2022 | Revised: 3 December 2022 | Accepted: 8 December 2022 | Published Online: 11 December 2022

Abstract

Science, technology, engineering, and mathematics (STEM) problem-solving is necessary to be infused into the classroom. Nevertheless, the criticism of underrepresented mathematics in STEM problem-solving assessment is an issue. In this study, we develop and investigate the psychometric evidence of an integrated STEM-based mathematical problem-solving test. The product of the test was a mathematical essay test that contains three scientific scenarios related to the environment in every middle school grade. The mathematical contents were integrated into engineering-based design using the technology. Three experts filled an assessment sheet to assess content validity, which was analyzed using a content validity index (CVI) and intraclass correlation coefficient (ICC). The result of content validity revealed that overall items were valid and reliable. The construct validity was examined using the Rasch analysis from the data of Grades 7-9 students in Indonesia (n = 286). The construct of all scenarios and prompting items indicated fit with various difficulty levels and acceptable discrimination value. Nevertheless, four prompting items were reported as misfit based on unweighted mean square value. The recommendation for improvement is emphasized in the language clarity aspect. The inter-rater reliability was also declared good. A further study is suggested to provide a computer-based test.

Keywords: Development, Mathematical Problem-Solving, Psychometric Evidence, STEM, Test

How to Cite: 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. http://doi.org/10.22342/jme.v13i4.pp587-604

Mathematics is a foundation for learning other disciplines and plays a pivotal role in our daily lives (Junpeng et al., 2020; Kesorn et al., 2020; Khairani & Sahari Nordin, 2011). However, graduates encounter difficulties in applying mathematics in real-life situations because practice-oriented mathematics education is isolated from the mathematics curriculum (Jones et al., 2015; Ke & M. Clark, 2020; Shute et al., 2016). This issue causes a shift in mathematics class to problem-solving mathematics in the real-world context (Jones et al., 2015). Problem-solving is the most significant cognitive activity in professional, every day, 21st century, and life-long learning (İncebacak & Ersoy, 2016; Karatas & Baki, 2013). Therefore, problem-solving in mathematics is a core standard of the mathematics curriculum in several countries (Du Toit & Du Toit, 2013; Karatas & Baki, 2013).

Problem-solving in mathematics includes procedure application, concept reasoning, synthesizing, analyzing, accessing information, and interpreting (Karatas & Baki, 2013). Infusing problem-solving in mathematics classes can improve mathematics achievement (Karatas & Baki, 2013). Several researchers have developed, validated, and examined students' problem-solving skills, particularly using closed-ended essay tests (Du Toit & Du Toit, 2013; İncebacak & Ersoy, 2016; Widodo et al., 2021). However, they used well-defined, monodisciplinary, and structured tests, which are fundamentally different from real-life problems (Shute et al., 2016). Hence, they encountered difficulties in both daily demands and international assessment of authentic problem solving (Shanta, 2019; Shute et al., 2016).

Advances in technological innovation and the economic industry globally have impacted the labor market (Kelley & Knowles, 2016; Maass et al., 2019). Professionals who can communicate and use information; solve issues in science, technology, engineering, and mathematics (STEM); adapt; and be innovative in response to changing demands and information are required (Maass et al., 2019). Moreover, global challenges, including social and environmental issues, increase rapidly (Jolly, 2016; Kelley & Knowles, 2016). Hence, the education area needs to improve students' knowledge in interdisciplinary by integrating STEM disciplines (Jolly, 2016; Kelley & Knowles, 2016; Maass et al., 2019). Because STEM is a complex concept, integrating its discipline requires problem solving (Maass et al., 2019).

Over the years, several researchers have developed frameworks related to STEM for formative assessment (Kelley & Knowles, 2016; Maass et al., 2019; Priemer et al., 2020; Wells, 2016). Researchers have also developed assessments using games or problem-based complex scenarios in STEM problem solving, mainly emphasizing science education (e.g., ALCHEMIST, Use Your Brainz, ChemLabBuilder) (Annaggar & Tiemann, 2020; Scherer & Tiemann, 2012; Shute et al., 2016). These assessments raised issues of the lacking role and underrepresentation of mathematics in STEM assessments (Lasa et al., 2020; Maass et al., 2019).

Another problem, teachers encounter difficulties in implementing STEM assessment due to a complex area of assessment, the lack of applicable resource, and practical problems (Amalina & Vidákovich, 2022b; Gao et al., 2020). Assessing students in STEM area is challenging since it needs to connect several disciplines according to the curriculum and framework, but in fact, there is a lack of explicitly and well-articulated framework in this area (Amalina & Vidákovich, 2022b). In addition, there are still lack of assessment tools that are applicable in the class (Amalina & Vidákovich, 2022b; Gao et al., 2020). They mostly research-oriented and time-consuming assessments. The constructed type test is the most applicable test for measuring cognitive in STEM education, but the psychometric evidence of the available tests are mostly difficult to obtain (Amalina & Vidákovich, 2022b).

Although integrating STEM in problem solving has been proven essential in middle school mathematics, the trend of STEM assessment is still monodisciplinary, specifically in the area of mathematical problem solving assessment in Indonesia (Amalina & Vidákovich, 2022b; Gao et al., 2020). However, the Indonesian 2013 mathematics curriculum requires students to master mathematical problem solving related to daily life and STEM integration. Consequently, students encounter difficulties in solving interdisciplinary problems (Suratno et al., 2020).

According to several issues raised by researchers, a test that emphasizes mathematical problem solving based on an integrated STEM framework for Indonesian middle school students should be developed. The test focuses on interdisciplinary knowledge and skills (cognitive) as a part of STEM activity assessment through problem solving. A mathematical task is situated in the written science context that requires students to apply their knowledge about mathematics, science, and engineering- based design using a specific tool to solve the problem. Therefore, in this study, we develop an integrated STEM-based mathematical problem-solving test, validate the content and construct of the developed test, and examine the reliability of the developed test using inter-rater reliability.

Framework in STEM Problem Solving

The definition of STEM education is an interdisciplinary learning approach that involves knowledge integration or combining two or more disciplines among science, technology, engineering and mathematics (Gao et al., 2020; Jolly, 2016). Science entails applying content, context, or a manner of thinking that provides a context for reflection, organization, and action. Mathematics provides a set of notions, skills, and contents for solving problems (Jolly, 2016; Lasa et al., 2020). Engineering offers an engineering-based design that enables the model construction (Lasa et al., 2020). Additionally, technology refers to technical tools that ease a process (Jolly, 2016; Lasa et al., 2020).

Infusing STEM approach in the education area is required since it includes the real-world context and problem solving (Maass et al., 2019; Tasir et al., 2018). Solving an interdisciplinary problem is a key component of STEM education that requires problem solving skills (Jolly, 2016). Problem solving provides a powerful context for learning and practicing STEM concepts, and draws on skills, approaches, and ways of thinking that can be applied across disciplines (Priemer et al., 2020). The process of solving STEM problem depends on the type of problem (Jolly, 2016). A problem that requires students to produce an artefact to demonstrate their mastery of content called problem-based activity (Jolly, 2016; Priemer et al., 2020). However, there is a type of STEM problem that requires students to apply their knowledge based on authentic STEM text or scenario (Shanta, 2019). An integrated STEM based-mathematical problem is not a problem-based activity. It is a mathematical problem in a science scenario that developed by integrating problem-solving framework in every discipline.

The available STEM problem solving frameworks are based on frameworks in every STEM discipline. In science, problem-solving is always embedded in the scientific inquiry framework and scientific reasoning (Priemer et al., 2020). Problem-solving in science emphasizes on inductive thinking, while in mathematics allows inductive and deductive thinking. Problem-solving in mathematics is a process to solve mathematical tasks that potential to provide intellectual challenge by understanding a problem and applying mathematical concepts to generate a conclusion (Lasa et al., 2020; Priemer et al., 2020). The preliminary mathematical problem-solving framework is heuristics that continuously followed by other framework related to it (Priemer et al., 2020). Researchers have exploited mathematics approaches closely related to scientific inquiry, such as mathematics experimentation and inquiry-based mathematics education. The opposite of the concept of "inquiry" is "proving" (Priemer et al., 2020). The problem solving framework in engineering is related to engineering design or creating a solution (Jolly, 2016; Kelley & Knowles, 2016; Wells, 2016).

The role of mathematics in STEM problem-solving is to provide a set of notions and skills that allow solving problems, including mathematical ability to analyze, reason, and interpret solutions (Jolly, 2016; Lasa et al., 2020). Another skill that has an important role in STEM problem-solving is mathematical modelling since STEM problem is a complex problem that requires to model from a real-world problem into a mathematical sentence (Maass et al., 2019). In addition, mathematics engages in STEM problem- solving by application of mathematics concepts and procedures to the problem (Lasa et al., 2020).

Several researchers also have developed STEM frameworks by applying frameworks in various disciplines, focusing on practical problem solving (Kelley & Knowles, 2016; Wells, 2016). Priemer et al. (2020) developed framework using a PISA framework as a ground framework and embedded some domain-specific problem-solving techniques, e.g., scientific inquiry, proving, engineering design, and computer science. Amalina and Vidákovich (2022a) developed a framework by modifying Priemer et al.'s (2020) framework to integrate STEM into mathematical problem-solving for summative and diagnostic assessments (see Table 1). We refer to Amalina and Vidákovich (2022a) framework to develop a test.

Table 1. An integrated STEM problem-solving framework

	Indicators
Exploring and understanding	(1) Identifying relevant and valuable unknowns and given information (STEM); (2) Determining the goal of a problem (STEM); (3) Employing a useful basic concept (SM)
Representing and formulating	(4) Constructing a tabular, graphical, symbolic, or verbal representation of a problem using technology (STM); (5) Making a hypothesis, developing criteria, or checking existing theory (STEM).
Planning and executing	(6) Formulating a reasonable argument, explanation, and solution (including strategy, step design, and model building) (TEM); (7) Arranging, critically choosing, and evaluating alternatives (STEM); (8) Performing a plan by deducting, proofing, or finding a counterexample, applying mathematics concepts, mathematization, reasoning, computational skills, science concept, and technology (STEM).
Monitoring and reflecting	(9) Drawing conclusions, evaluating, and reflecting on results and methods (STEM).

Note: S = science, T = technology, E = engineering, M = mathematics

Content and Context in STEM Education

Based on the analyses of the Cambridge curriculum, Indonesian curriculum, PISA framework, and literature review, mathematics contents used in STEM tasks are mostly measurements, geometry, and arithmetic. Moreover, the contexts used are mainly scientific contexts related to the environment and how to protect it. These contents and contexts can be integrated into STEM problem solving and real-world problems (Lasa et al., 2020). There are several content knowledge intersections in every grade.

1. Mathematics: Arithmetic (number, fraction, and measurement), Algebra (ratio and proportion), Geometry (two-dimensional (2D) figure), and Statistics (data representation and central tendency)
2. Science: Pollution, climate and disaster, eco-friendly product, electricity, and energy.

There is no specific STEM curriculum in Indonesia. Science and mathematics are taught separately at middle school (grades 7 to 9) or International Standard Classification of Education (ISCED) level 2. Technology serves as a tool in the teaching and learning process. Nevertheless, the importance of integrating STEM in the Indonesian mathematics classroom has been emphasized.

Although there are no differences in grades' main competencies, basic competencies may be different and higher in some grades. Some basic competencies in a certain grade can become the prerequisites for basic competencies in a higher grade. Problem-solving competencies are the compulsory competencies for applying mathematics concepts. Solving problems related to numbers and fractions are compulsory basic competencies in every grade. In Grade 7, students have to master problem solving in ratios, proportions, 2D figures, and data representations. Students in Grade 8 have to acquire knowledge of solving 2D and three-dimensional (3D) figures and central tendency problems. Grade 9 students should master the prior knowledge of function (represent data), 3D figure (including prior knowledge about circle), and congruency and similarity (ratio and proportion).

METHOD

Instrumentation

The instrument we used is an integrated STEM-based mathematical problem-solving test, a developed complex scenario essay test. There are three scientific contexts in every grade. Every scientific context has eight items (prompting items) to explore students' problem-solving skills using indicators in Table 1. The scoring used in every prompting item is from 0 (blank answer) to 5 (complete and correct answer). Students are allowed to use any tool or application to help them. The time given in every scenario is 1 h.

The tool for assessing content validity is an assessment sheet that experts will complete. It assesses scenarios, items, and scoring. It contains four aspects: sufficiency, clarity, coherence, and relevance aspects. The sufficiency aspect includes the suitability of aims, curriculum, indicators, and student level. The clarity aspect measures the language (syntax and semantics). The coherence aspect refers to the relatedness to dimension and indicator. The relevance aspect measures the importance level of an item. Experts must rate using four scales (from 1 = not clear to 4 = very clear). The qualitative assessment is also provided by giving suggestions on each item for improvement.

Procedure

The procedures for developing the test are: (1) analyze the research needs and develop a framework, (2) analyze the contents and contexts that could be applied in the test, (3) develop scenarios of the test related curriculum contents and contexts, (4) develop prompting items in every scenario based on the problem-solving indicators and basic competencies, and (5) generate the answer and scoring method.

After developing the instrument, content validation was performed by expert panels. The experts rated the instrument using the assessment sheet. The experts were recruited with criteria: (1) a minimum of a master's degree in mathematics education, (2) must be an educator, and (3) a minimum of three years of working experience. The qualitative assessment from experts in each item will be used to improve items. After revising based on the experts' suggestions, the test was administered to middle school students to analyze the construct validity. The data collection was performed between March 22 and April 12, 2022. The test was administered in 3 h using Google Forms.

Because the test is an essay, we need to ensure consistency in the implementation of a rating system. As such, two raters–one of the authors and a female teacher who has a master's degree in mathematics education–checked students' answers. The students' scores from two raters were compared and analyzed using statistical tests.

Participants

Three female experts were selected to assess content validity. They are secondary school mathematics teachers in an urban area who have a master's degree in mathematics education. The test was conducted on Grades 7-9 students in East Java, Indonesia, to check the psychometric evidence. We selected A-accreditation schools with random classes. Five schools participated, with a total of 286 participants from three areas. Table 2 summarizes the characteristics of the participants.

Table 2. Characteristics of participants

Demographic	Characteristics	N	%
Gender	Boys	102	35.66
	Girls	184	64.34
Grade	7 (M age = 13.09, SD = 0.61)	116	40.56
	8 (M age = 13.99, SD = 0.49)	90	31.47
	9 (M age = 14.95, SD = 0.48)	80	27.97

School location	City	128	44.76
	District	158	55.24
Ethnic	Javanese	279	97.55
	Madurese	4	1.40
	Others	3	1.05

Note: N = the number of students, SD = Standard deviation, M age = mean age

Data Analysis

Content validity was analyzed using the content validity index (CVI) and intraclass correlation coefficient (ICC). The CVI is a quantitative evaluation to measure the mean content validity ratio given by experts. The CVI is the most widely used index for content validation. The ICC is the correlation test to measure agreement among non-binomial scores given by experts. Since in the current study there are three experts rated the test by 1-4, the ICC is applicable to measure how similar are the scores given by experts.

The construct validity was examined using Rasch analysis, which described (1) the item fit model, (2) reliability, (3) item discrimination, (4) the item level of difficulties, (4) the Wright map, and (6) scale step analysis. The consistency of implementation of the rating scale (inter-rater reliability) was examined using ICC. The analysis was performed using SPSS 25 and Conquest applications.

The model used in the Conquest application is the partial credit model. The fit of this model was proven significantly better than the fit of the rating scale model according to the Conquest manual book. To explore the item fit model, we used the weighted (infit) mean square (MNSQ) and unweighted (outfit) MNSQ values. The ideal value of outfit and infit MNSQs is 1 based on the Rasch measurement model, but items are categorized as fit if the outfit or infit MNSQ values of [0.5, 1.6) (Andrich, 2018; Bond & Fox, 2015). The acceptable item discrimination value is more than 0.2.

A separation reliability coefficient evaluates whether the localization parameters of the items are sufficiently separate to cover the entire ability interval. The parameters are separate enough if the reliability coefficient is greater than or equal to 0.90 or $\chi 2$ is significant (López-Pina et al., 2016). The expected a posteriori/plausible value (EAP/PV) reliability is an estimate for test reliability provided by the Conquest software.

The ability parameters of the students are compared with the item localization parameters through a Wright map, which also shows the item-level difficulty. Moreover, the item-level difficulty is reported through estimate values. The higher the estimated values, the more difficult the item is. The scale step analysis aims to analyze the scoring function or behavior. The scale behavior is represented by the "Pt bis" value–the additional result of a category and correlation of those items– and the percentage of students who achieve every scale (0-5).

We applied Rasch analysis for both prompting items and scenarios. Scenarios act as items that construct the test, and a prompting item refers to a subitem of the test. Hence, it is necessary to examine the extent to which the test accurately assesses problem-solving skills through the scenarios and prompting items. We used the total score of students answering the prompting items in every scenario to analyze the fit model in every scenario. The total score is 40, but we recode it for the Conquest application. We recode them into a 1-5 scoring scale with the scoring rubric. We performed a Rasch analysis using the students' scores in every prompting item to analyze the fit model of prompting items. There are eight prompting items in every scenario. The total number of scenarios is six. Thus, there are 48 prompting items in total. The score used is from 1 to 5 in each prompting item.

RESULTS AND DISCUSSION

Developing an Integrated STEM-Based Mathematical Problem-Solving Test

The assessment trend in STEM is an essay scenario problem test that we used as a basis to develop a new test. The integrated STEM-based mathematical problem-solving test is a problem-based complex scenario test that integrates mathematics through the new STEM framework (Table 1). It is a mathematical scenario word problem related to scientific phenomena and applies engineering design to solve a problem with the help of technology. Mathematics is taking parts as contents, notions, and skills in problem solving. Science participates in the manner of thinking and the context used. Engineering engages in engineering-based design, and technology provides a tool used during problem solving.

There are three scenarios in every grade related to environmental management with one or two anchor scenarios in every grade. Every scenario has a challenge. The contents used are number and measurement, ratio and proportion, geometry, and statistics. Each scenario has eight dependent prompting items. The prompting items are developed from the sub indicators, integration between the indicators (Table 1) and basic competencies of middle school Indonesian 2013 curriculum. The prompting items in every topic are metacognitive prompting questions for guiding students to explore their problem-solving steps and reveal their thinking. The use of technology as aid is required (in this case, calculators, grid papers, rulers, and any possible devices are allowed).

The maximum score in each indicator is 5, adapted from Docktor et al. (2016) and Salazar-Torres et al. (2021). Score 5 for a complete and correct answer, 4 for a complete answer with minor error, 3 for incomplete but correct or complete with major error, 2 for incomplete answer with minor error, 1 for completely wrong and irrelevant or incomplete with major error, and 0 for a blank answer. The total maximum score for every scenario is 40. Table 3 explains the description of each scenario.

Table 3. Description of every scenario in the integrated STEM-based mathematical problem-solving test

Scenario	Grade	Contents	Descriptions
Eco-Friendly Packaging	7	Area (square and rectangle), simple arithmetic calculation, and decimal	Design eco-friendly packaging with simple constraints and decide on the lowest price
School Park	7	Ratio, area and circumference of rectangle, diameter, fraction, designing graph, and simple arithmetic calculation	Design a planted school park with the trees that can absorb the largest amount of CO_2 and represent the CO_2 absorbed in several years by graph
Calory versus Greenhouse Gas Emission	7, 8, and 9	Percentage, decimal, and simple arithmetic calculation	Design a 1-day-menu that fulfills the nutrition and has the lowest CO_2 emission during their productions
Flood Water Reservoir	8	Central tendency, table, simple arithmetic calculation, and volume	Design a flood water reservoir and decide the fastest time by pumps to absorb the average volume
City Park	8 and 9	Percentage, ratio, decimal, designing graph, diameter, complex arithmetic calculation, conversion, and rectangle area	Design a planted city park with the trees that can absorb the largest amount of CO_2 and represent the remaining CO_2 emission in several years by graph after planting
Infiltration Well	9	Diameter, ratio, circle, and complex arithmetic calculation	Design an infiltration well with complex constraints and decide the price to build it

The contents tested in "Calory versus Greenhouse Gas Emission" are the contents that the students already studied in Grade 7 and are basic competencies (as well as basic competencies of prior knowledge) in all grades. Therefore, the scenario could be tested for all grades. The contents in "City park" were already studied in Grades 7-9; however, because of complex constraints, they will be administered in Grades 8 and 9. The "Infiltration Well" contents are only suitable for Grade 9 because the cylinder (and the prior knowledge: circle) topic appears in the Grade 9 curriculum. Moreover, the "Flood Water Reservoir" could not be administered in Grade 7 because the central tendency topic appears in Grade 8. Figure 1 illustrates an example of a scenario in the test and the prompting items.

Scenario 1: Eco-Friendly Packaging

Eco-friendly and low-budget products are starting to be prioritized. We want to produce packaging with the following criteria.

1. The packaging has length, width, and height of 18 cm, 18 cm, and 8 cm, respectively
2. The cover of the packaging is square and 1.2 times the area of the base.
3. There are four material options:

	Duplex paper	Ivory paper	Styrofoam	Plastic
Size	79 cm × 109 cm	215 mm × 330 mm	100 cm × 50 cm	17 cm × 50 cm
Thickness	250 gsm	250 gsm	0.5 cm	0.5 cm
Price	Rp 4,000.00	Rp 2,000.00	Rp 9,000.00	Rp 1,300.00
Time for decomposing	2-6 months		Cannot decompose	1,000 years

Challenge

Design an eco-friendly packaging and decide on the lowest budget for packaging a product based on the criteria.

Prompting questions:

1. What is your challenge (asked) in the scenario?
2. What information do you need to resolve the challenge?
3. What is the total area of four sides, cover, and base in cm^2?
4. What is your guess relating to the type of material that could be used to design the packaging? Give your reasons.
5. Draw your packaging design based on the shape and length.
6. Mention two materials that are unsuitable to be chosen and give your reasons!
7. How much is the price for your packaging based on your chosen material and area of your packaging?
8. Is the packaging eco-friendly and the cheapest one? Give your conclusion about the price and type of material used to produce packaging.

Figure 1. Example of a scenario and prompting items in the developed test

Content Validity of the Developed Test

The CVI value for every scenario is 1. However, all experts agreed with a score of 4 (very clear) for only the "Eco-Friendly Packaging." In the "Calory versus CO2 Gas Emission" and "Infiltration Well" scenarios, all experts agreed with a score of 3 (clear) due to the need for minor revision regarding language clarity and sufficiency at the student level. The "City Park" and "School Park" scenarios received one 4 score and two 3 scores due to several items' problems. Moreover, in the "Flood Water Reservoir" scenario, due to language barrier, only an expert scored 3. See Table 4 for detailed numerical results.

Table 4. The content validity score for every scenario

Scenario	Expert 1's score	Expert 2's score	Expert 3's score
Eco-friendly packaging	4	4	4
School park	4	3	3
Calory versus gas emission	3	3	3
Flood water reservoir	4	4	3
City park	4	3	3
Infiltration well	3	3	3

The CVI values of the prompting items in every scenario ranged between 0.67 and 1 (see Table 5). An item with a CVI of 0.67 is an item that measured students' skills in generating a conclusion in every scenario. However, only one judge gave a score of 2, and two judges granted a score of 3. Judges stated that the item is somewhat relevant, but another item may be covering what this item is measuring.

Table 5. The content validity score for every prompting items

Expert 1	Expert 2	Expert 3	CVI	Prompting items
4	4	4	1	1, 3–7, 9, 11–15, 17, 19–23, 25, 27–31, 33, 35–39
3	3	4	1	10, 41–47
3	4	4	1	2, 18
3	2	3	0.67	8, 16, 24, 32, 40, 48
4	3	4	1	26, 34

The experts agreed that the scenarios and prompting items suffice to measure student level based on curriculum and indicators and were related to the measuring dimension or indicator. However, several terms and sentences needed minor revision. The Cronbach alpha indicated acceptable with .835, and the ICC result represented moderate reliability (rxx = 0.628).

Qualitative recommendations are suggested by judges: (1) reduce the item and the hints in the prompting items, (2) change the scenarios' format, and (3) make a uniform score. The previous part's description of the test development is the revision version after the experts' assessment.

Construct Validity of the Developed Test

General Analysis Based on Scenario

The results of Rasch analysis showed 52 iterations, 27 estimated parameters, and 1,800.663 final deviance. The weighted MNSQ value ranged between 0.84 and 1.12, with various scenario difficulty levels. Scenario 2 was the most difficult (estimate value = 1.307) and scenario 4 was the easiest (estimate value = −1.141). Table 6 describes the response model parameter estimates in every scenario.

Table 6. Response model parameter estimates based on scenario

Scenario	Estimate	Error	Unweighted MNSQ	Weighted MNSQ
1. Eco-Friendly Packaging	0.610	0.138	0.86	0.89
2. School Park	1.307	0.135	0.86	0.84
3. Calory versus Gas Emission	−0.156	0.106	1.22	1.12
4. Flood Water Reservoir	−1.141	1.148	1.10	0.93
5. City Park	−0.986	0.123	0.98	0.94
6. Infiltration Well	0.366	0.293	0.96	0.96

Rasch analysis also measured the difficulty level for students to make an achievement transition from one step to another. In this case, the score used in every scenario is from 0 to 5. Hence, the Rasch analysis calculated the transition among these scores. The results that revealed the hardest transition in all scenarios was from score 4 to 5 and the easiest transition was from score 1 to 2 or 0 to 1, indicating that students found it easier to reach score 2 when they had score 1 and students found it more difficult to reach score 5 when they had score 4. Scenario 3 had the most difficult transition from score 4 to 5 compared with other scenarios, with an 8.968 estimate value. Additionally, scenario 3 had the easiest transition from score 0 to 1 compared with the others, with a −7.572 estimate value. Figure 2 (a) depicts the Wright map based on the scenario.

The discrimination values of the scenarios ranged between 0.85 and 0.94. Pt bis was constantly increasing, and toward the last values, it was positive. For instance, the Pt bis values of scores 0-5 in scenario 5 are −0.49, −0.37, −0.37, −0.31, 0.12, and 0.59, respectively. The percentage of students who achieved every scale clarified the equal distribution, indicating that the behavior of the scoring scale is acceptable and fit. The separation reliability was high at 0.984 (χ^2 (5)=239.341,p<.001), and the EAP/PV reliability was 0.867. The high separation reliability indicates that all scenarios can cover all students' abilities. Furthermore, the high EAP/PV reliability shows that the test is reliable.

Detailed Analysis Based on Prompting Items

Rasch analysis performed 158 iterations, 224 total estimated parameters, and a final deviation of 14,130.920. The weighted MNSQ values ranged between 0.66 and 1.53. Four prompting items had unweighted MNSQ values of more than 1.60, namely, items 4, 19, 36, and 38. Nevertheless, the weighted MNSQ values of all prompting items indicated fit. Item 4, regarding making hypotheses in the "Eco-Friendly Packaging" scenario, had an unweighted MNSQ value of 2.63. Item 19, about using the basic concept in the "Calory versus Gas Emission" scenario, had a 1.82 unweighted MNSQ value. Items 36 and 38 request students to make a hypothesis and arrange critical choices in the "City Park" scenario, which had 1.62 and 1.92 unweighted MNSQ values, respectively.

The difficulty levels were various, with item 25 as the easiest prompting item (estimate value = −1.766) and item 47 as the most difficult prompting item (estimate value = 2.065). Item 25 requires students to determine the goal in the "Flood Water Reservoir" scenario. Meanwhile, item 47 requests students to apply the concept in the "Infiltration Well" scenario.

Regarding the difficulty levels for achievement transition from one score to another, the results revealed that the most difficult levels for students are to achieve from score 3 to 4 and score 4 to 5. Seven prompting items (Items 4-6, 11, 12, 36, and 46) reported that the most difficult transition of student achievement was from score 3 to 4. The 41 other prompting items reported that the most difficult transition of student achievement was from score 4 to 5. The hardest transition was prompting Item 21 from score 4 to 5 (estimate value = 6.333).

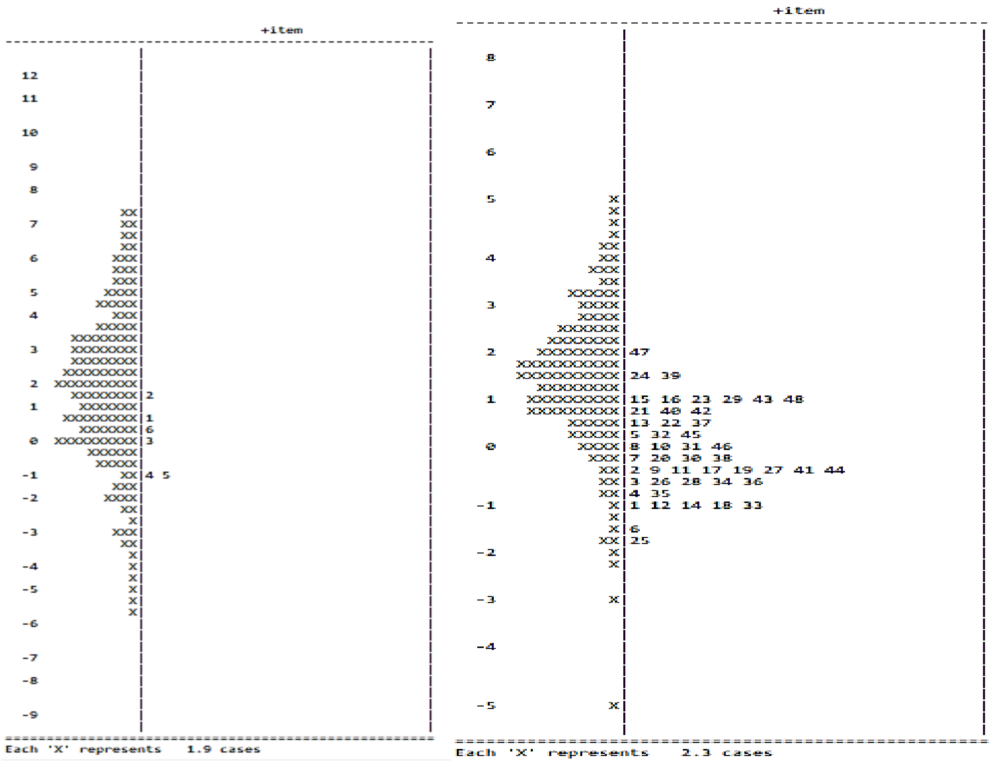
The easiest achievements were from score 0 to 1, score 1 to 2, and score 2 to 3. Four items reported that the easiest transition was from score 2 to 3, namely, items 41-43 and 47. Twelve prompting items (items 1, 2, 9, 10, 17, 26, 34, 40, 44-46, and 48) notified that the easiest transition was from score 1 to 2. The other prompting items showed that the student achievement transition from 0 to 1 was the easiest. The easiest transition was reported on item 21 from score 0 to 1 (estimate value = −6.539). Figure 2 (b) explains the Wright map of the distribution of student achievement based on items.

The discrimination values of the items ranged between 0.64 and 0.87. Pt bis was constantly increasing, and toward the last values, it was positive. The percentage of students who achieved every scale clarified equal distribution. The separation reliability was high at 0.986 (χ^2 (47)=3,411.843,p<.001) and the EAP/PV reliability was 0.962. The high separation reliability indicates that all prompting items can cover all students' abilities. Furthermore, the high EAP/PV reliability shows that the test is reliable.

Furthermore, we discuss the misfit items. The items are misfit because the unweighted MNSQ value is more than the expected value. The unweighted MNSQ value is less important since Conquest analyzed the observed curve (curve from empirical data) based on the weighted MNSQ results compared with the model curve that has an MNSQ value of 1. According to the results, the observed curves of these misfit items were a little flatter than the model curve (underfit). This will often be the case when the MNSQ is significantly greater than 1, indicating that the data were less predictable than the model expected or the data had more variation in the observed pattern response. For instance, the weighted MNSQ for item

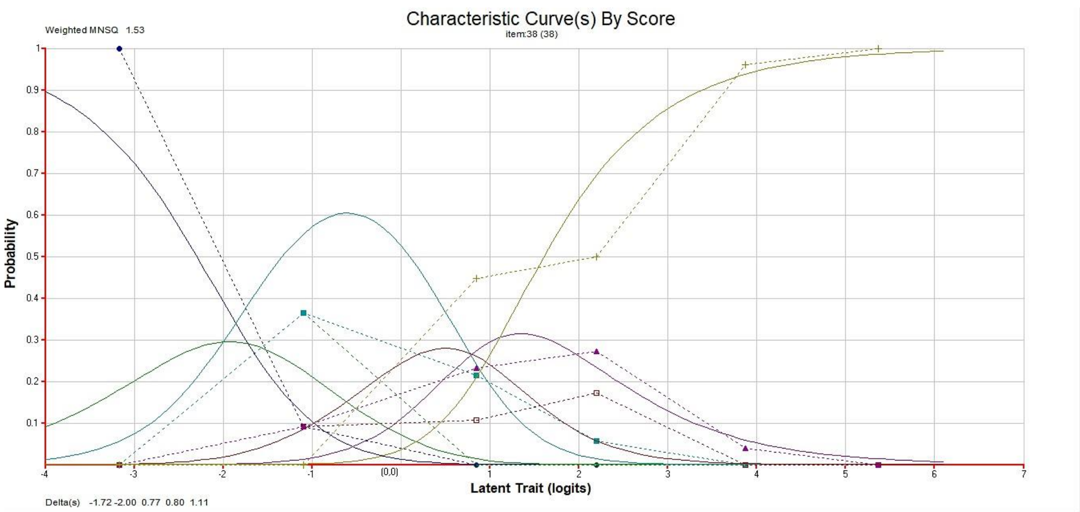
38 was 1.53, indicating that there was 53% more randomness (i.e., noise) in the data than modeled. Hence, further investigation is required to determine the reason.

Figure 3 shows the characteristic curve for each category (five scores) for item 38. Empirical and modeled curves are matched via color. The disparity between the observed and modeled curve for score 2 and 3 were the largest, and this was consistent with the high fit statistics for this category (weighted MNSQ value more than 1.6), meaning that students who received a score of 2 and 3 behaved differently when answering Item 38.



Note for scenario: Item 1: Eco-Friendly Packaging; Item 2: School Park; Item 3: Calory versus Gas Emission; Item 4: Flood Water Reservoir; Item 5: City Park; Item 6: Infiltration Well.

Figure 2. (a) Item (scenario) and students' ability distribution map. (b) Item (prompting item) students' ability distribution map.



Note: Scores 1-5: blue, green, light blue, red, and light brown, respectively.

Figure 3. Item characteristic curve for Item 38

Similar patterns were found for items 4, 19, and 36. Students who received a score of 2 and 3 behaved differently. It was observed from the curve and confirmed from the highest fit statistics compared with other categories. The weighted MNSQ values for score 2 and 3 in items 4, 19, and 36 were above 1.6.

Reliability of the Integrated STEM-Based Mathematical Problem-Solving Test

Students' answers (n = 286) to the test were checked by two experts. The mean score of every student from the two experts was analyzed using ICC. The results revealed a good ICC (rxx = 0.958) and good reliability (α = 0.979). According to the result, the consistency of the implementation of a rating scale was proven.

Discussion

Assessment in interdisciplinary problem solving, specifically in STEM area, is required to tackle 21st-century challenges. However, issues related to underrepresented mathematics, the trend of monodisciplinary problem solving in Indonesia, and the lack of diagnostic tests in this area have been detected (Lasa et al., 2020; Maass et al., 2019). Hence, the development of an integrated STEM-based mathematical problem-solving test resolves these issues.

The contents and contexts of the test were developed based on the Indonesian 2013 curriculum. An essay scenario-based problem was chosen as the type of test. The essay test is suitable to be applied to assess students' problem-solving, which is process-oriented rather than result-oriented (Gao et al., 2020). Scenario-based problem is the most common type of test in interdisciplinary problem solving because it embodies the authenticity of a problem that engages students in it (Gao et al., 2020; Shanta, 2019; Tasir et al., 2018). The anchor scenario was raised from the intersection of basic competencies that appear between grades to compare the problem-solving achievement of Grades 7-9.

Metacognitive skills are implied and practiced during mathematical problem solving. Metacognition in problem solving focuses on planning, monitoring, evaluating, and reflection (Velasquez & Bueno, 2019). Metacognitive skills cannot be assessed explicitly. Hence, metacognitive prompt questions are required. Metacognitive prompting items in the developed test are designed to reveal student thinking, ensure student responses were consistent, address the students' abilities identified, and explore every problem-solving indicator. It agrees with the study aim of Shanta (2019) that uses metacognitive prompts in his developed test.

We performed content validity, construct validity, and reliability assessments to ensure the psychometric appropriateness of the developed test. Practitioners were chosen to assess content validity because they understand the real condition of the students and curriculum. The result of ICC indicated moderate (Koo & Li, 2016). The CVI was adequate for all scenarios, but several prompting items had a CVI value of less than 0.69. CVI is categorized as adequate if it is greater than 0.69; however, if 50% of the judges gave a score of 3 or 4, a CVI value of less than 0.69 is still acceptable (Hyrkäs et al., 2003). In this case, the prompting items that had a CVI value of 0.67 were generated from two scores of 3 and one score of 2. Because more than 50% of judges (two of three) gave a score of 3, the CVI values of these prompting items were acceptable. The low CVI value was because of the few number of judges.

The result of construct validity revealed all scenarios were categorized as fit based on chosen theories (Andrich, 2018; Bond & Fox, 2015). The discrimination items, behavior of the score, and reliabilities were acceptable (Bond & Fox, 2015). The most difficult scenario was the "School Park." The most common difficulty was that students could not differentiate between area and circumference concepts that they want to use, which agrees with the previous research that there is misalignment between geometry item difficulty and students' abilities (Bostic & Sondergeld, 2015). Moreover, students failed to make hypotheses and decide the best option from several options. Providing several options perhaps increases student difficulty in solving the scenario. The easiest scenario was "Flood Water Reservoir" because of the simple constraint of the available options.

The achievement transition from score 4 to 5 was the most difficult because score 5 requires a perfect answer. The easiest transition is from score 0 to 1 because score 0 is for a blank answer and score 1 is for an irrelevant answer or wrong answer. It was rare for students in all phases of problem solving to receive a score of 0. Hence, the probability of receiving more than a score of 0 is high.

A detailed analysis was performed to check the model fit of prompting items. According to the weighted MNSQ results, all prompting items were fit (Andrich, 2018; Bond & Fox, 2015). Conquest analyzed the observed curve based on the weighted MNSQ results compared with the model curve. Hence, fit items can be detected only with weighted MNSQ, but it will be more accurate to observe both weighted and unweighted MNSQs. However, the unweighted MNSQ results for four prompting items were misfit. The unweighted MNSQ is more sensitive to responses to items with difficulty far from a person or vice versa (Bond & Fox, 2015). These are usually easy to remedy and diagnose (e.g., lucky guess or careless mistake).

According to item characteristic curves, the data from students who received a score of 2 and 3 caused misfits in these items. We decided not to delete these items because it will distort the test's construct. Items 4, 36, and 38 requested students to make hypotheses and eliminate impossible answers. After checking the students' answers, mostly they were confused between "possible" and "impossible" words as well as "guess or hypotheses" words. Hence, they put "possible guess" into the "impossible guess." However, based on their general answers, they understood how to make hypotheses and evaluate alternatives. Therefore, it can be concluded that the misfit items were because of the language barrier. Item 19 is about using the basic concept used in solving a problem, in this case, calculating the required protein, fat, and carbohydrate. The common mistakes were (1) students only mentioned the required nutrition (in percentage) based on the given information without calculating them, and (2) some students interpreted the word "fulfill" in the question with "more than," but the rest interpreted it as "less than." However, after examining their complete problem-solving process, they understood how to calculate them. They misinterpreted the question. This is perhaps because of an ambiguous prompting question. The original prompting item was "How many kcal of fat, carbohydrate, and protein are needed to fulfill the criteria?" and it was revised to "calculate the fat, carbohydrate, and protein needed based on the criteria in kcal." Additionally, we modified the word "fulfill" into "more than."

The discrimination items, behavior of the score, and reliabilities in the prompting items were acceptable (Bond & Fox, 2015). The most difficult item was item 47, which requested the students to calculate the price to build an infiltration well. The most common error was that students could not differentiate between the concept of surface area and volume that they wanted to use, which agrees with the result of Tan Sisman and Aksu (2016). Additionally, item 47 requires complex skills and concepts, e.g., number and fraction, 3D figure, ratio and proportion, and circle. Item 25 is the easiest because it only needs students to understand the already stated question's goal. The difficulty levels for transiting achievement in the prompting items were similar to the difficulty levels for transiting achievement in the scenario, with the transition from score 4 to 5 as the most difficult and 0 to 1 as the easiest transition.

Inter-rater reliability is necessary for measuring the consistency of the implementation of a rating score in essay tests. The reliability was certified as good (Koo & Li, 2016).

CONCLUSION

The development of a mathematical problem-solving test based on an integrated STEM framework addressed the issues of previous studies with emphasis on mathematics while still ensuring interdisciplinarity. Six essay scenario-based problems with anchor scenarios in every grade were developed. Metacognitive prompting items were provided in every scenario based on indicators. The developed test was proven to be adequate based on experts' judgments and content assessment.

The construct validity assessment was performed using the Rasch model, which indicated that all scenarios were fit with a reasonable difficulty level. The behavior of the score, EAP/PV reliability, and discriminant value were acceptable. Additionally, the weighted MNSQ value, discriminant value, and EAP/PV reliability of all prompting items denoted fit and acceptable. However, four prompting items showed misfit according to unweighted MNSQ values.

The suggested revision of misfit items was emphasized in the language aspect. The suggested revision is formulated according to the result of the item characteristic curve and the most common errors encountered by students.

This study has limitations regarding the test and the sampling method. The developed test was only administered using Google Forms because of the lack of students' knowledge of other platforms. Additionally, the roles of engineering and technology in the test are underrepresented compared to mathematics and science. Regarding the sample, it is suitable to involve students from different schools (e.g., sample Grade 8 was only from a school). Only involving students from a school limits the external validity.

The results of the current study will be beneficial for mathematics teachers in assessing students mathematical problem-solving skills in an interdisciplinary context. Additionally, the test can be used as well for science teachers for assessing specific science topics. Teachers can use the test as an assessment in interdisciplinary knowledge and skills as a part of STEM activity assessment. This test can be used as a preliminary test before conducting an action or experiment in a mini laboratory to prompt students' knowledge of an interdisciplinary STEM topic. Hence, before practicing in an activity, students have a plan regarding what they want to do and interdisciplinary understanding of the issue related to environmental management. The test emphasized on process rather than the product itself. Teachers can extend the test into practice to assess a product based on what students planned in the test.

For researchers in the relevant area, the results of the test validity can be a basis for adapting the test in a different sample background and for further testing of its validity. The future study is widely open as well regarding providing the test in a computer-based format.

Declarations

Author Contribution : IKA: Conceptualization, Writing, Formal analysis, Methodology, Editing and Visualization.
TV: Review, Validation and Supervision

Funding Statement : This article received a funding from the Doctoral School of Education, University of Szeged, Hungary.

Conflict of Interest : No potential conflict of interest was reported by the authors.

Additional information : Additional information regarding the test is available from the first author

REFERENCES

Amalina, I. K., & Vidákovich, T. (2022a). A mathematical problem-solving framework-based Integrated

STEM: Theory and practice. *International Journal of Trends in Mathematics Education Research*, *5*(1), 1-11. https://doi.org/10.33122/ijtmer.v5i1.105

Amalina, I. K., & Vidákovich, T. (2022b). Assessment in STEM Problem-Solving: A Systematic Review. *The International Journal of Assessment and Evaluation*, *29*(2), 64-80. https://doi.org/https://doi.org/10.18848/2327-7920/CGP/v29i02/63-80

Andrich, D. (2018). Advances in social measurement: A Rasch measurement theory. In F. Guillemin, A. Leplège, A. Briançon, E. Spitz, & J. Coste (Eds.), *Perceived health and adaptation in chronic disease* (pp. 66-91). https://psycnet.apa.org/doi/10.1201/9781315155074-7

Annaggar, A., & Tiemann, R. (2020). A video game as an assessment tool of problem-solving competence. *Gesellschaft Fur Didactik Der Chemie Und Physik (GDCP)*, *April*, 377-380.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model : Fundamental Measurement in the Human Sciences Third Edition*. Routledge.

Bostic, J. D., & Sondergeld, T. A. (2015). Measuring Sixth-Grade Students' Problem Solving: Validating an Instrument Addressing the Mathematics Common Core. *School Science and Mathematics*, *115*(6), 281-291. https://doi.org/10.1111/ssm.12130

Docktor, J. L., Dornfeld, J., Frodermann, E., Heller, K., Hsu, L., Jackson, K. A., Mason, A., Ryan, Q. X., & Yang, J. (2016). Assessing student written problem solutions: A problem-solving rubric with application to introductory physics. *Physical Review Physics Education Research*, *12*(1), 1-18. https://doi.org/10.1103/PhysRevPhysEducRes.12.010130

Du Toit, S., & Du Toit, G. (2013). Learner metacognition and mathematics achievement during problem-solving in a mathematics classroom. *The Journal for Transdisciplinary Research in Southern Africa*, *9*(3), 505-518. https://doi.org/10.4102/td.v9i3.194

Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education*, *7*(1). https://doi.org/10.1186/s40594- 020-00225-4

Hyrkäs, K., Appelqvist-Schmidlechner, K., & Oksa, L. (2003). Validating an instrument for clinical supervision using an expert panel. *International Journal of Nursing Studies*, *40*(6), 619-625. https://doi.org/10.1016/S0020-7489(03)00036-1

İncebacak, B. B., & Ersoy, E. (2016). Problem Solving Skills of Secondary School Students. *China-USA Business Review*, *15*(6), 275-285. https://doi.org/10.17265/1537-1514/2016.06.002

Jolly, A. (2016). *STEM by design: Strategies and activities for grades 4-8*. Routledge. https://doi.org/10.4324/9781315679976

Jones, I., Swan, M., & Pollitt, A. (2015). Assessing Mathematical Problem Solving Using Comparative Judgement. *International Journal of Science and Mathematics Education*, *13*(1), 151-177. https://doi.org/10.1007/s10763-013-9497-6

Junpeng, P., Marwiang, M., Chiajunthuk, S., Suwannatrai, P., Chanayota, K., Pongboriboon, K., Tang, K. N., & Wilson, M. (2020). Validation of a digital tool for diagnosing mathematical proficiency. *International Journal of Evaluation and Research in Education*, *9*(3), 665-674. https://doi.org/10.11591/ijere.v9i3.20503

Karatas, I., & Baki, A. (2013). The Effect of Learning Environments Based on Problem Solving on Students ' Achievements of Problem Solving. *International Electronic Journal of Elementary Education*, *5*(3), 249-267. https://www.iejee.com/index.php/IEJEE/article/view/25

Ke, F., & M. Clark, K. (2020). Game-Based Multimodal Representations and Mathematical Problem Solving. *International Journal of Science and Mathematics Education*, *18*(1), 103-122. https://doi.org/10.1007/s10763-018-9938-3

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1). https://doi.org/10.1186/s40594-016-0046-z

Kesorn, N., Junpeng, P., Marwiang, M., Pongboriboon, K., Tang, K. N., Bathia, S., & Wilson, M. (2020). Development of an assessment tool for mathematical reading, analytical thinking and mathematical writing. *International Journal of Evaluation and Research in Education*, *9*(4), 955-962. https://doi.org/10.11591/ijere.v9i4.20505

Khairani, A. Z., & Sahari Nordin, M. (2011). The Development and Construct Validation of the Mathematics Proficiency Test for 14-Year-Old Students. *Asia Pacific Journal of Educators and Education*, *26*(1), 33-50. http://web.usm.my/apjee/webtest/APJEE_26.1.2011/26.1.2011_33-50.pdf

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Lasa, A., Abaurrea, J., & Iribas, H. (2020). MATHEMATICAL CONTENT on STEM ACTIVITIES. *Journal on Mathematics Education*, *11*(3), 333-346. https://doi.org/10.22342/JME.11.3.11327.333-346

López-Pina, J. A., Meseguer-Henarejos, A. B., Gascón-Cánovas, J. J., Navarro-Villalba, D. J., Sinclair, V. G., & Wallston, K. A. (2016). Measurement properties of the brief resilient coping scale in patients with systemic lupus erythematosus using rasch analysis. *Health and Quality of Life Outcomes*, *14*(1), 1-8. https://doi.org/10.1186/s12955-016-0534-3

Maass, K., Geiger, V., Ariza, M. R., & Goos, M. (2019). The Role of Mathematics in interdisciplinary STEM education. *ZDM - Mathematics Education*, *51*(6), 869-884. https://doi.org/10.1007/s11858-019-01100-5

Priemer, B., Eilerts, K., Filler, A., Pinkwart, N., Rösken-Winter, B., Tiemann, R., & Zu Belzen, A. U. (2020). A framework to foster problem-solving in STEM and computing education. *Research in Science and Technological Education*, *38*(1), 105-130. https://doi.org/10.1080/02635143.2019.1600490

Salazar-Torres, J., Rincón Leal, O., & Vergel Ortega, M. (2021). The rubric as an assessment tool for solving problem situations in the physics and mathematics teaching context. *Journal of Physics: Conference Series*, *1981*(1). https://doi.org/10.1088/1742-6596/1981/1/012018

Scherer, R., & Tiemann, R. (2012). Factors of problem-solving competency in a virtual chemistry environment: The role of metacognitive knowledge about strategies. *Computers and Education*, *59*(4), 1199-1214. https://doi.org/10.1016/j.compedu.2012.05.020

Shanta, S. (2019). Developing and assessing authentic problem-solving skills in high school pre-engineering students. *ASEE Annual Conference and Exposition, Conference Proceedings*. https://doi.org/10.18260/1-2--32637

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106-117. https://doi.org/10.1016/j.chb.2016.05.047

Suratno, Wahono, B., Chang, C. Y., Retnowati, A., & Yushardi, Y. (2020). Exploring a Direct Relationship between Students' Problem-Solving Abilities and Academic Achievement: A STEM Education at a Coffee Plantation Area. *Journal of Turkish Science Education*, *17*(2), 211-224. https://doi.org/10.36681/tused.2020.22

Tan Sisman, G., & Aksu, M. (2016). A Study on Sixth Grade Students' Misconceptions and Errors in Spatial Measurement: Length, Area, and Volume. *International Journal of Science and Mathematics Education*, *14*(7), 1293-1319. https://doi.org/10.1007/s10763-015-9642-5

Tasir, Z., Mohamad, M. N. M., Ismail, Z., & Mustafa, N. (2018). Development and Validation of Problem Solving Task Based-Integrated STEM. *Proceedings - 2018 6th International Conference on Learning and Teaching in Computing and Engineering, LaTiCE 2018*, 17-23. https://doi.org/10.1109/LaTICE.2018.00011

Velasquez, Q. F., & Bueno, D. C. (2019). Metacognitive Skills in Problem Solving of Senior High School STEM Strand Students Metacognitive Skills in Problem-Solving among Senior High School STEM Strand Students. *Institutional Multidisciplinary Research and Development*, *2*(June), 124-129. https://doi.org/10.13140/RG.2.2.18820.99207

Wells, J. G. (2016). PIRPOSAL model of integrative STEM education: Conceptual and pedagogical framework for classroom implementation. *Technology and Engineering Teacher*, *75*(6), 12-19. http://america.aljazeera.com/articles/2013/9/16/many-us-bridges-arestructurallyunsoundsaysnewreport.html

Widodo, S. A., Ibrahim, I., Hidayat, W., Maarif, S., & Sulistyowati, F. (2021). Development of Mathematical Problem Solving Tests on Geometry for Junior High School Students. *Jurnal Elemen*, *7*(1), 221-231. https://doi.org/10.29408/jel.v7i1.2973

Study 3: Assessment of domain-specific prior knowledge: a development and validation of mathematical problem-solving test.

(Amalina & Vidákovich, 2023a)

Assessment of domain-specific prior knowledge: A development and validation of mathematical problem-solving test

Ijtihadi Kamilia Amalina[1], Tibor Vidákovich[2]
[1]Doctoral School of Education, University of Szeged, Szeged, Hungary
[2]Institute of Education, University of Szeged, Szeged, Hungary

Article Info

Article history:

Received Jan 31, 2022
Revised Nov 3, 2022
Accepted Nov 24, 2022

Keywords:

Domain-specific prior knowledge
Mathematical problem-solving
Psychometric analysis
Test

ABSTRACT

Science, technology, engineering, and mathematics (STEM) is a complex problem-solving that depends on the deep structures of domain-specific prior knowledge (DSPK) in mathematics. However, there is a lack of mathematics DSPK tests measuring several mathematics topics in every problem-solving phase in conceptual and procedural knowledge. This study aims to develop a mathematical problem-solving test as a mathematics DSPK test and investigate the content and construct validity. The product of test development is a 30-multiple-choice-item test in six mathematics topics. Every topic underlined all problem-solving phases in conceptual and procedural knowledge in a science or individual context. There were six experts performed the content validity sheets which analyzed using the content validity index (CVI) and intraclass correlation coefficient (ICC). The construct validity was examined using the Rasch model from 175 data of 7th grader students in Indonesia (Mage=12.66, SD=55). The result of content validity revealed overall items were valid (CVI≥83) and reliable (α=863; rxx=513). The construct of all items indicated fit (90≤weigheted MNSQ≤1.16) and were reliable (α=74) with various levels of difficulties and six low discrimination items. The recommendation for improvement is emphasized in language aspects. The absence of knowledge of facts could be an improvement for further study.

Corresponding Author:

Ijtihadi Kamilia Amalina
Doctoral School of Education, University of Szeged
6722 Szeged, Petőfi S. sgt.32-34, Hungary
Email: ijtihadi.kamilia.amalina@edu.u-szeged.hu

1. INTRODUCTION

Problem-solving is the goal of education and part of the science and mathematics curriculum [1]. The higher-order thinking and problem-solving model in the monodisciplinary area were developed in the 1960s' with the assumption by learning problem-solving, a student could transfer to any situation in daily life [1], [2]. Problem-solving is a complex set of cognitive, behavioral, and attitude components seen as a situational and context bond process that depends on the deep structures of knowledge and experiences [2], [3].

The development of problem-solving research is based on the theory that people encounter complex problems in their life, which require transdisciplinary [4], [5]. Moreover, the development of technology and economic globalization requires students to be good at transdisciplinary problem-solving in science, technology, engineering, and mathematics [5]. Hence, science, technology, engineering, and mathematics (STEM) problem-solving is infused into education. Since STEM problem-solving involves complex components in several disciplines, STEM problem-solving demands mastering domain-specific prior knowledge (DSPK) [6].

Several cognitive factors influencing problem-solving are prior knowledge, knowledge-based, formal reasoning ability, long-term memory, working memory, neo-Piagetian, and metacognitive variables [1], [2]. The most crucial factor is domain-specific knowledge under the theory of prior knowledge [2], [6], [7]. Prior knowledge is a stock of information, skills, beliefs, and experiences located before the learning process [8]. Other prior knowledge terms are pre-storage, permanent storage knowledge, background knowledge, pre-existing knowledge, and domain-specific knowledge [8]. Domain-specific knowledge engages in organizing new knowledge, encoding and representing information, assimilating new material, selecting relevant information to be elaborated, recalling, and retrieving information [2], [8]–[10].

In STEM problem-solving, it is essential to investigate students' DSPK in mathematics because mathematics plays a main part in the concept and content applications. However, there is a lack of study regarding measuring DSPK or developing assessment DSPK tools in mathematics. Several researchers focused on physics and biology [11]–[14], chemistry [15]–[17], and health education [18]. They implemented multiple-choice test [11], [18], two-tiered instrument [17], open-ended essay, complex essay, and map concept task [12], [13]. According to these issues, developing a mathematics DSPK instrument is necessary.

DSPK positively impacts problem-solving skills through experimental and non-experimental design research. Studies in DSPK in teaching and learning for primary to undergraduate students revealed that activating DSPK affects success in complex and creative problem-solving as well as self-regulation [1], [3], [4]. Non-experimental studies were conducted to find the correlation and impact of DSPK in mathematics on problem-solving. DSPK in mathematics was proven to impact students' complex and creative problem-solving skills [6], [19]. Moreover, studies about mathematics DSPK, specifically declarative (conceptual) and procedural knowledge, correlated and affected mathematical problem-solving [7], [20]–[23]. As a matter of fact, mathematics DSPK accounted for 56% of the variance [7], and conceptual understanding explained 31% of the variance in problem-solving [21].

DSPK in mathematics could be declarative (conceptual), procedural, strategic, and situational knowledge [2], [4], [12], [20]. Mostly, the researchers conducted procedural and conceptual research by applying essays (open-ended and short answers) and multiple-choice tests [4], [20]–[23]. However, they only focused on applying a phase in a topic rather than on all problem-solving phases in several mathematics topics. Moreover, DSPK assessment tools should concern the knowledge researchers want to assess in their main test. Hence, they could correlate with each other and measure the related knowledge. Consequently, not all previous assessment tools in mathematic DSPK could be adapted. According to the importance of mathematics DSPK (both procedural and conceptual knowledge) in complex problem-solving, in this case, STEM problem-solving, and the need for assessment tools on it, hence we discuss some issues in this paper regarding i) How is the development of the mathematical problem-solving test to measure DSPK in STEM problem-solving?; ii) How is the content validity of the developed mathematical problem-solving test?; and iii) How is the construct validity of the developed mathematical problem-solving test?

2. THEORETICAL BACKGROUND

2.1. Domain-specific prior knowledge, mathematical problem-solving, and their assessment tools

Prior knowledge is categorized as content knowledge and metacognitive knowledge [24]. Content knowledge is knowledge about domain or subject-oriented knowledge that contains procedural and conceptual (declarative) knowledge. Metacognitive knowledge refers to what a person knows about themself. It is divided into task knowledge, strategic knowledge, and self-knowledge.

Procedural knowledge is knowledge about how and when to use the appropriate skills and procedures [4], [20], [21]. It is the knowledge that contains actions and manipulations that are valid within a domain [2]. Conceptual knowledge, known as declarative knowledge in some resources, is knowledge about what and why a procedure is appropriate for a task, including checking the reasonable of the procedure [21]. It is knowledge about concepts, facts, definitions, operations, relations, and principles [2], [20]. Declarative knowledge could be assessed using multiple-choice and short answers, procedural knowledge is measured mainly by performance tests and possible by using multiple-choice and short answers, and strategic knowledge is rarely to be measured directly [2].

Hailikari, Nevgi, and Ylänne [24] distinguished declarative and procedural knowledge based on the cognitive domain. Declarative knowledge is divided into the knowledge of fact (KF) and knowledge of meaning (KM). The lower level of declarative knowledge has a low level of abstraction with simple recognition (i.e., recognition, enumerating, recalling, and remembering). The higher level of declarative knowledge needs deeper and requires the ability to define the concept (i.e., defining, reproducing, and understanding the meaning of the concept). Procedural knowledge is divided into the integration of knowledge (IK) and the application of knowledge (AK). The low level of procedural knowledge requires an ability to see the interrelation between concepts and how different phenomena link, including classifying and comparing. The higher level involves applying knowledge, performing a problem-solving task, producing, and implementing.

The types of knowledge in [24] engage in heuristic mathematical problem-solving phases, namely understanding a plan, devising the plan, carrying out the plan, and looking back. KM emphasizes the understanding phase, while KF does not include it in the problem-solving phase due to low abstraction knowledge. IK focuses on understanding, planning, and evaluating phases, while AK is concerned with applying phase.

Studies regarding the positive effects of DSPK assessments in complex and creative problem-solving were conducted using multiple-choice and constructed item tests [4], [6], [19]. The multiple-choice test focused on the relationship between variable and action knowledge (declarative and procedural knowledge) concerning the decision-making test [4] and conceptual knowledge [6]. DSPK related to conceptual (declarative) and procedural knowledge impacted mathematical problem-solving measured by constructed and selected types of tests [7], [20], [21], [25]. Some studies focused on algebraic problem solving (including proportion) and the prior knowledge related to it (i.e., fraction and proportion) [25] and algebra [21]. The basic concept knowledge and problem representation were also assessed by using short reasoning answer test [21]. A similar test was applied by [25] and concluded that conceptual knowledge about fractions is the most correlated factor (r=45) influencing proportional problem-solving.

2.2. Misconceptions, difficulties, and error in mathematics

Error in mathematics is caused by misconception, carelessness, reading problems, and some other factors. A misconception is different from an error since a misconception is a lack of understanding of a mathematics concept or rule [26]. A misconception about mathematics basic concepts was still found. There were 33.9% of Indonesian students had difficulties ordering arithmetic calculations, operating the whole number, and understanding negative signs [27]. In addition, students have difficulties converting length, weight, and time because they lack experience in their daily life [26].

The one-step higher concept is a fraction that is important to master higher concepts in algebra and statistics. However, students encountered misconceptions related to the meaning of the fraction (e.g., the meaning of fraction as a part of a whole, the meaning of fraction as a quantity (a quotient relation between two numbers) rather than two separate whole numbers). Moreover, they had problems adding fractions with the whole number, converting a fraction into a different representation (e.g., decimal, percentage), and calculating decimals because of the place value [26]–[28]. The related topic with the fraction is proportion. It was found that students had difficulties in conceptual, procedural, and factual knowledge [29]. They could not understand the concept of direct and indirect proportion. Hence, they could not distinguish between them. Regarding geometry topics, students had problems with both conceptual and procedural. They could not differentiate between the formula of area or circumference and volume or surface area, served 3D figures as 2D figures, and had a misconception about units [30].

The most serious difficulty in mathematical problem-solving is understanding the problem. Researchers interviewed primary teachers and concluded that students had difficulties in: i) Understanding keywords appearing in the problem; ii) Figuring out what to assume and what information is necessary; and iii) Reading and motivation problems [31]. When students have difficulties in the first phase of problem-solving, they cannot go to the higher phases (i.e., applying and evaluating).

3. RESEARCH METHOD

3.1. Instrumentation

The test is a 30-multiple-choice curriculum-based knowledge test. It measures students' conceptual and procedural knowledge of mathematical problem-solving. The topics used are numbers (integer and fraction) & measurement, ratio & proportion, geometry, and statistics. Each topic consisted of understanding (conceptual knowledge), planning, applying, and evaluating (procedural knowledge) phases in a science or individual context. There are four options for each question. The total score is 30 in a 90-minutes test. They completed the test online and could not use any helping device (e.g., a calculator). At the end of the test, there are demographic questions and a space for adding comments related to the drawback of the instrument (time, number of items, media, and difficulty).

The tool for assessing content validity is an assessment sheet that experts will complete. It contains three aspects: content, construct, and language aspects. The content aspect includes the suitability of aims, curriculum, indicators, and students' level. The construct aspect refers to the construction of questions, options, and information. The language aspect measures both clarity and ambiguity. Experts have to rate with 4 scales (e.g., 1=not clear and 4=very clear). The qualitative assessment is also provided by giving suggestions on every item for improvement.

3.2. Procedure

The procedure for developing the mathematical problem-solving test consisted of three main phases. The first phase analyzes the curriculum, students' needs, and topics we want to use in the test. The second phase generates the test indicators. The third phase generates the questions and options based on the indicators.

After developing the instrument, content validation was performed by expert panels. The experts' panels rated the instrument by using the assessment sheet. The expert panels are recruited based on several criteria: i) A minimum of a master's degree in mathematics or mathematics education; ii) Educators; and iii) A minimum of three years of working experience. The qualitative assessment from experts in each item will be used to identify and improve items that were mathematically inaccurate, ambiguous, or item prompting student responses that did not indicate their understanding of conceptual and procedural knowledge in mathematical problem-solving. It was validated after revising a draft by administering it to secondary school students. The objective of this phase is to verify that the test had good construct validity and reliability.

3.3. Participants

Six experts were selected to assess content validity, a male and five females. Of those, two lecturers in mathematics education and four mathematics teachers (three graduated from mathematics education, and one from mathematics). They graduated from different universities and currently work in different institutes in urban areas. The participants for investigating construct validity were 175 7th grader students from public schools in East Java, Indonesia (Mage=12.66, SD=55). Of those, 85 male and 90 female with 98.35% identified as Javanese, 1.1% were Madurese, and 6% were Sundanese.

3.4. Data analysis

The content validity was analyzed using the content validity index (CVI) and intraclass correlation coefficient (ICC). The construct validity was examined by using Rasch analysis. It described the item and person's behavior, including the item fit model, discrimination, item difficulties, and behavior of options. The analysis was performed with SPSS 25 and Conquest applications.

4. RESULTS AND DISCUSSION

4.1. Developing mathematical problem-solving test

STEM problem-solving requires complex skills, knowledge, cognitive components, and disciplines [2], [3]. Hence, it is needed to assess mathematics domain-specific knowledge in STEM problem-solving [8]. The conceptual and procedural knowledge in DSPK is important to be assessed because it influences problem-solving skills [7], [20]–[23]. Moreover, most Indonesian students had difficulties with basic concepts of mathematics. Hence, based on needs analysis, it requires developing a test for assessing DSPK in conceptual and procedural knowledge. Conceptual knowledge is divided into the knowledge of fact (KF) and knowledge of meaning (KM), and procedural knowledge is categorized into the integration of knowledge (IK) and the application of knowledge (AK) [24]. Since knowledge of fact has a low level of abstraction, it is not included in the development of the mathematical problem-solving test.

A multiple-choice test is the most appropriate type of test to measure DSPK [2], [4], [6]. Hence, we develop a multiple-choice test with four options. The options for every item have distractions. The distractions are the consequence of choosing the wrong answer in the previous number (or phase), some common misconceptions, common errors, and difficulties that students encounter. The misconceptions, difficulties, and common errors are: i) Interpreting text [31]; ii) Understanding the meaning of fractions [26]; iii) Converting a fraction into decimal and percentage [27]; iv) Ordering calculation/properties of number [27]; v) Understanding the decimal point/place value [26], [28]; vi) Figuring out what to assume, what is asked, what information from the problem is necessary to solve [31]; vii) Understanding conversion [26]; viii) Understanding and differentiating direct and inverse proportions [29]; ix) Differentiating between area and perimeter [30].

The test is targeted at junior secondary students (7th-9th graders), but we validated it on 7th graders only. The topics used are the topics applied in the main test and intersection topics for all graders based on the Indonesian curriculum and literature review. The topics are number (integer and fraction (n=9)) & measurement (n=4), ratio & proportion (n=8), geometry (n=4), and statistics (n=5). Each topic consisted of understanding, planning, applying, and evaluating phases in a science or individual context. KM measures students understanding phase of mathematics basic concepts (e.g., the meaning of fraction concepts). IK targets interrelation among concepts in understanding, planning, and evaluating phases (e.g., finding the correct statement regarding their problem-solving process). AK focuses on the application of knowledge in applying phase (e.g., calculating the reduction of CO2 emission). Figure 1 explains the example of the test.

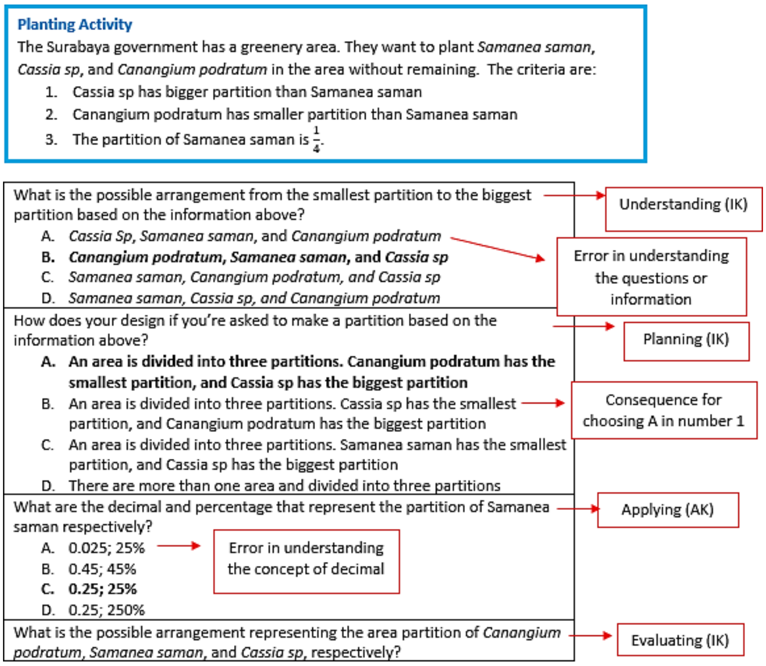


Figure 1. Example of the test in number topic

The indicators are from the Indonesian curriculum, mathematical problem-solving framework, and are related to the mathematics concept used in the main test. Table 1 explains the outline of the item for the mathematical problem-solving test. The total score is 30 in a 90-minutes test. The total time for conducting the test, answering demographic questions, and adding comments related to the developed test is 120 minutes.

Table 1. The outline of the item for the mathematical problem-solving test.

Basic competence	Item
Explain the concept of fractions and decide the order of integers and fraction	3
Solve problems related to ordering integers and fraction	1, 2, 5
Explain and calculate integers and fractions (and an approximation) by applying their properties	4, 8, 12
Solve problems related to integer calculation and a fraction (as well as an approximation) by applying integer properties	6, 7, 9, 10, 11, 13
Explain the concept of the ratio	18
Differentiate the concept of direct and inverse proportions by (or from) table, graph, and equation	19
Solve problems related to direct and inverse proportions by using a table, graph, or equation	14, 15, 16, 17, 20, 21
Solve contextual problems related to the area and perimeter/ circumference of the 2D figure	22, 23, 24, 25
Represent and solve a problem related to distribution, mean, median, modus, and the spread of data for making decisions or prediction	26, 27, 28, 29, 30

4.2. Content validity of the mathematical problem-solving test

There were two items (items 10 and 18) with a CVI of 83, an item with a CVI of 67 (item 11), and the rest with a perfect CVI value. Items 10 and 18 had a problem in the content aspect with a CVI of 83. These items were considered as a high level of difficulty. Item 11 had a problem in both content and language aspects with a CVI of 67, and it was categorized as an inadequate CVI [32]. Experts argued that item 11 had a high level of difficulty and ambiguity. Since item 11 was categorized as inadequate, it needs to be revised. Experts suggested providing more needed information and supplying new wording because of the high item difficulty.

The experts presented several suggestions mostly related to language aspects, including: i) Adjusting students' language ability; ii) Revising the ambiguous question; and iii) Supplying new wording to clarify the question. The ambiguous questions make readers misunderstand, leading them to different or double answers. The item that lacks information could produce a double answer as well. For example, an item requires students to find the length and width of a rectangle with an area of 20 cm^2, respectively. The options are "4 and 5" or "10 and 2". The researcher wants students to choose "10 and 2"; hence, it is needed to add "l>w" to clarify.

Regarding the content, they met an agreement that the curriculum, indicator, and topics were appropriate for junior secondary students. However, several items were predicted to have a high level of difficulty. The given recommendation was to reduce the constraints or add valuable information. The evaluation phase items require students to decide on a true statement based on the given information. It is a high-difficulty item and requires students to reading comprehension. Therefore, re-phrase and supplying useful information were necessary. It was also recommended to add constraints in the low-level difficulty items (e.g., in the items related to numbers and measurements) and add relevant information in the high-level difficulty items (e.g., in the items related to geometry and statistics). The reliability of the mathematical problem-solving test was good, with a Cronbach Alpha of .863. Moreover, the intraclass correlation coefficient indicated moderate reliability with 51 [33].

4.3. Construct validity of the mathematical problem-solving test

The Rasch analysis was applied to investigate the construct validity with 60 iterations, 31 estimated parameters, and 6401.500 final deviance. All 30 items were categorized as fit, with MNSQ values ranging between 90 and 1.16. The value of weighted MNSQ is considered fit if it ranges between 6-1.4 [34]. The test could cover mostly all student abilities. However, several students with below-average abilities still could not solve the easiest item. Based on this theory of Rasch, it is recommended to add easier items or change the participants to a higher grade. Since it was a pilot study administered to 7th graders, the test targets 7th-9th grade students. Hence, it would be appropriate for them in the real test (the difficulty of the test will be appropriate for 8-9 grade students). The item difficulty level was distributed from the easiest level in item 1 (estimate value of -1.063) into the hardest level in item 30 (estimate value of 1.246). Item 1 had the easiest level because it only measures the simple understanding of given information with low constraints. Item 30 had the hardest level because it measures students' ability in evaluating phase, which needs higher-order thinking skills and complex calculations.

The item separation reliability indicates how well the item parameters are separated. In this case, the separation reliability was high with 941. The coefficient alpha value indicated the test was reliable at 74, and the EAP/PV reliability was 706. Figure 2 describes response model parameter estimates and Figure 3 represents a map of latent distributions and response model parameter estimates.

There were six items with low discrimination values, namely items 10 (discrimination of 14), 11 (discrimination of 08), 16 (discrimination of 08), 21 (discrimination of 16), 27 (discrimination of 17), and 29 (discrimination of 08). However, those items' MNSQ values were categorized as fit, and the options behaviors were acceptable. The addition result of a category (option) and correlation of those items represented acceptable value. The key answer had the highest value, except item 11. However, in items 6, 21, and 29, the probabilities of high-ability students clicking the different options from the key answers were higher, but the key answers still had the highest voter. After checking their previous related items, they clicked the right answers. Hence, their behavior in these items could be because of guessing or error. It could also occur because the options of these items consisted of a high level of distraction (seems similar option) that directed students to choose them.

VARIABLES				UNWEIGHTED FIT			WEIGHTED FIT		
	item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1	1	-1.063	0.129	0.88	(0.79, 1.21)	-1.2	0.92	(0.90, 1.10)	-1.6
2	2	-0.857	0.128	0.96	(0.79, 1.21)	-0.4	0.97	(0.91, 1.09)	-0.7
3	3	-0.655	0.128	1.05	(0.79, 1.21)	0.5	1.05	(0.91, 1.09)	1.0
4	4	-0.832	0.128	0.97	(0.79, 1.21)	-0.3	0.99	(0.91, 1.09)	-0.3
5	5	-0.094	0.130	1.01	(0.79, 1.21)	0.1	1.02	(0.89, 1.11)	0.4
6	6	0.323	0.135	0.96	(0.79, 1.21)	-0.3	0.93	(0.85, 1.15)	-1.0
7	7	0.095	0.132	0.96	(0.79, 1.21)	-0.4	0.94	(0.88, 1.12)	-0.9
8	8	-0.555	0.128	0.88	(0.79, 1.21)	-1.1	0.90	(0.91, 1.09)	-2.3
9	9	0.538	0.138	0.86	(0.79, 1.21)	-1.4	0.90	(0.83, 1.17)	-1.1
10	10	0.040	0.131	1.13	(0.79, 1.21)	1.2	1.10	(0.88, 1.12)	1.7
11	11	0.670	0.140	1.19	(0.79, 1.21)	1.7	1.14	(0.81, 1.19)	1.4
12	12	0.444	0.136	0.89	(0.79, 1.21)	-1.0	0.91	(0.84, 1.16)	-1.1
13	13	0.264	0.134	1.12	(0.79, 1.21)	1.2	1.09	(0.86, 1.14)	1.3
14	14	-0.605	0.128	1.05	(0.79, 1.21)	0.5	1.03	(0.91, 1.09)	0.7
15	15	-0.680	0.128	0.98	(0.79, 1.21)	-0.2	0.98	(0.91, 1.09)	-0.5
16	16	0.382	0.135	1.22	(0.79, 1.21)	2.0	1.15	(0.85, 1.15)	1.9
17	17	-0.580	0.128	1.00	(0.79, 1.21)	0.0	1.02	(0.91, 1.09)	0.4
18	18	-0.832	0.128	0.97	(0.79, 1.21)	-0.3	0.97	(0.91, 1.09)	-0.6
19	19	0.635	0.140	0.90	(0.79, 1.21)	-0.9	0.94	(0.82, 1.18)	-0.6
20	20	-0.404	0.128	0.89	(0.79, 1.21)	-1.0	0.91	(0.91, 1.09)	-2.1
21	21	0.263	0.134	1.16	(0.79, 1.21)	1.4	1.10	(0.86, 1.14)	1.4
22	22	0.121	0.132	0.95	(0.79, 1.21)	-0.4	0.96	(0.87, 1.13)	-0.6
23	23	0.066	0.131	1.03	(0.79, 1.21)	0.4	1.00	(0.88, 1.12)	0.1
24	24	-0.122	0.130	1.12	(0.79, 1.21)	1.1	1.07	(0.89, 1.11)	1.3
25	25	0.845	0.144	0.92	(0.79, 1.21)	-0.7	0.93	(0.79, 1.21)	-0.7
26	26	0.234	0.133	0.91	(0.79, 1.21)	-0.9	0.90	(0.86, 1.14)	-1.4
27	27	0.066	0.131	1.15	(0.79, 1.21)	1.4	1.10	(0.88, 1.12)	1.6
28	28	0.635	0.140	0.94	(0.79, 1.21)	-0.5	1.00	(0.82, 1.18)	-0.0
29	29	0.412	0.136	1.22	(0.79, 1.21)	1.9	1.16	(0.85, 1.15)	2.0
30	30	1.246*	0.714	0.96	(0.79, 1.21)	-0.4	1.01	(0.73, 1.27)	0.1

An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability = 0.941
Chi-square test of parameter equality = 484.323, df = 29, Sig Level = 0.000
^ Quick standard errors have been used

Figure 2. Response model parameter estimates

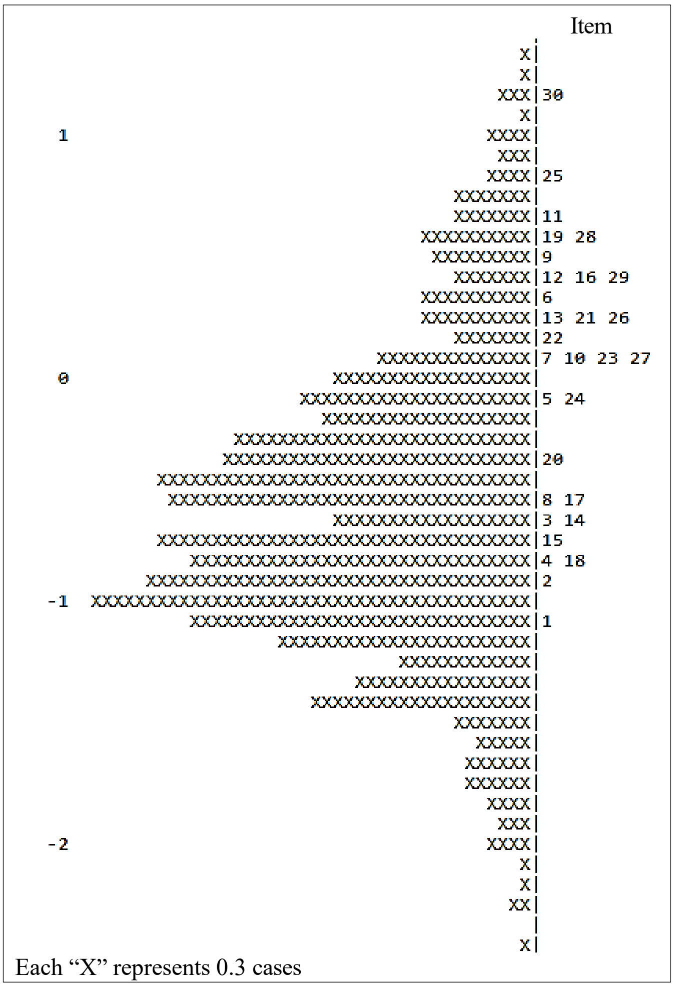


Figure 3. Map of latent distributions and response model parameter estimates

The behavior of distractors in item 11 was indicated as different from the theory. Most students clicked the C option (25.71%) rather than the right option, which is B (24%). The probability of students with a high ability level choosing option C was higher than choosing option B as shown in Figure 4. It was because of the language barrier. In item 11, the words "vertical" and "horizontal" made students encounter difficulty. Since the MNSQ value of item 11 was acceptable, we decided to revise rather than delete it [34].

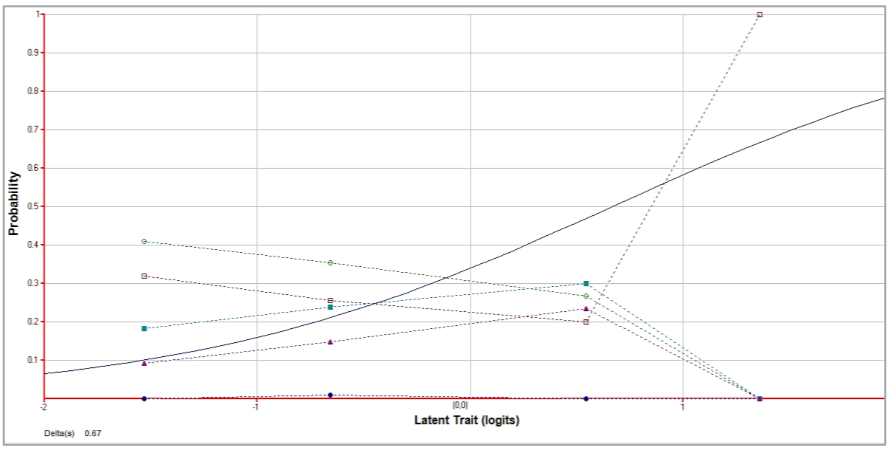


Figure 4. The characteristic curve of item 11 by category

Results of students' evaluation of the instrument revealed that 36% of students argued that they had a problem with the time given to solve the problem. We considered the time for solving the test to be from the Indonesian national examination, which is three minutes for every multiple-choice item. Regarding the test difficulty, 68% of students stated that the test was difficult. The number of items might influence the test difficulty. There were 40% of students expressed a problem with the high number of items. Since the test was multiple-choice, the probability of guessing is wide. A test with many items is required to avoid this issue. Moreover, some items and options need reading comprehension. There were 33% of students indicated that they had a problem with language understanding. Some items had been revised regarding language barriers. Regarding the media used, because the test is computer-based, there is a possibility for technical problems 6.3% of students encountered a technical problem during the test.

The test covered several contents and required higher-order thinking. Hence, the absence of knowledge of a fact becomes a drawback. Moreover, the number of items that measure knowledge of meaning, integration of meaning, integration of knowledge, and application of knowledge was not balanced. However, every phase of problem-solving in every content had occurred. The test was a curriculum-oriented test targeted at 7th-9th graders students. Hence it only could measure Indonesian junior secondary school students. In addition, the construct validity should be analyzed from sample grades 7th-9th, which was missing in this study.

5. CONCLUSION

The development of the problem-solving test serves as a DSPK test. It covers the limitation of previous studies by providing conceptual and procedural knowledge and every phase of problem-solving. It also provides a single assessment tool to measure students' DSPK in several mathematics contents. The product was a 30-item of multiple-choice test on six mathematics topics covered in a scientific or individual context.

The validation is a part of instrument development, including content and constructs validities. The mathematical problem-solving test had a good content validity index and reliability. Moreover, the construct validity indicated all items were fit. However, several items needed consideration and revision regarding the language barrier. The final product of the mathematical problem-solving test could be administered on a large scale to measure Indonesian junior secondary students' mathematical DSPK in complex problem-solving.

REFERENCES

[1] K. Serafino and T. Cicchelli, "Cognitive theories, prior knowledge, and anchored instruction on mathematical problem solving and transfer," *Education and Urban Society*, vol. 36, no. 1, pp. 79–93, Nov. 2003, doi: 10.1177/0013124503257016.

[2] J. J. Solaz-Portolés and V. S. López, "Cognitive variables in science problem solving: A review of research," *Journal of Physics Teacher Education Online*, vol. 4, no. 2, pp. 25–32, 2007.

[3] B. Rittle-Johnson, J. R. Star, and K. Durkin, "The Importance of Prior Knowledge When Comparing Examples: Influences on Conceptual and Procedural Knowledge of Equation Solving," *Journal of Educational Psychology*, vol. 101, no. 4, pp. 836–852, 2009, doi: 10.1037/a0016026.

[4] H. M. Süß and A. Kretzschmar, "Impact of cognitive abilities and prior knowledge on complex problem solving performance - Empirical results and a plea for ecologically valid microworlds," *Frontiers in Psychology*, vol. 9, no. MAY, pp. 1–22, 2018, doi: 10.3389/fpsyg.2018.00626.

[5] T. R. Kelley and J. G. Knowles, "A conceptual framework for integrated STEM education," *International Journal of STEM Education*, vol. 3, no. 1, 2016, doi: 10.1186/s40594-016-0046-z.

[6] C. Y. Chang, "Does problem solving = prior knowledge + reasoning skills in earth science? An exploratory study," *Research in Science Education*, vol. 40, no. 2, pp. 103–116, 2010, doi: 10.1007/s11165-008-9102-0.

[7] C. Oksuz, "Association of domain-specific knowledge and analytical ability with insight problem solving in mathematics," *International Journal of Pedagogies and Learning*, vol. 5, no. 1, pp. 138–153, Jan. 2009, doi: 10.5172/ijpl.5.1.138.

[8] T. Hailikari, *Assessing university students' prior knowledge. Implications for theory and practice*. Helsinki University Print, 2009.

[9] N. M. Crooks and M. W. Alibali, "Noticing relevant problem features: Activating prior knowledge affects problem solving by guiding encoding," *Frontiers in Psychology*, vol. 4, pp. 1–10, 2013, doi: 10.3389/fpsyg.2013.00884.

[10] D. C. Geary, D. B. Berch, R. J. Ochsendorf, and K. M. Koepke, *Acquisition of complex arithmetic skills and higher-order mathematics concepts*. Elsevier Inc., 2017.

[11] J. Wang, A. Stebbins, and R. E. Ferdig, "Examining the effects of students' self-efficacy and prior knowledge on learning and visual behavior in a physics game," *Computers and Education*, vol. 178, p. 104405, 2022, doi: 10.1016/j.compedu.2021.104405.

[12] T. Binder, A. Sandmann, B. Sures, G. Friege, H. Theyssen, and P. Schmiemann, "Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression," *International Journal of STEM Education*, vol. 6, no. 1, 2019, doi: 10.1186/s40594-019-0189-9.

[13] T. Binder, P. Schmiemann, and H. Theyssen, "Knowledge acquisition of biology and physics university students—the role of prior knowledge," *Education Sciences*, vol. 9, no. 4, pp. 1–18, 2019, doi: 10.3390/educsci9040281.

[14] S. Kantahan, P. Junpeng, S. Punturat, K. N. Tang, P. Gochyyev, and M. Wilson, "Designing and verifying a tool for diagnosing scientific misconceptions in genetics topic," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 9, no. 3, pp. 564–571, 2020, doi: 10.11591/ijere.v9i3.20544.

[15] O. Gulacar, A. Milkey, and S. Mclane, "Exploring the effect of prior knowledge and gender on undergraduate students' knowledge structures in chemistry," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 15, no. 8, 2019, doi: 10.29333/ejmste/106231.

[16] I. G. Ionas, D. Cernusca, and H. L. Collier, "Prior knowledge influence on self-explanation effectiveness when solving problems : An exploratory study in science learning," *International Journal of Teaching and Learning in Higher Education*, vol. 24, no. 3, pp. 349–358, 2012.

[17] A. Versprille *et al.*, "Assessing student knowledge of chemistry and climate science concepts associated with climate change: Resources to inform Teaching and learning," *Journal of Chemical Education*, vol. 94, no. 4, pp. 407–417, 2017, doi: 10.1021/acs.jchemed.6b00759.

[18] M. E. Beier and P. L. Ackerman, "Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment," *Psychology and Aging*, vol. 20, no. 2, pp. 341–355, 2005, doi: 10.1037/0882-7974.20.2.341.

[19] G. Molnár, S. Greiff, and B. Csapó, "Inductive reasoning, domain specific and complex problem solving: Relations and development," *Thinking Skills and Creativity*, vol. 9, pp. 35–45, 2013, doi: 10.1016/j.tsc.2013.03.002.

[20] M. A. Al-Mutawah, R. Thomas, A. Eid, E. Y. Mahmoud, and M. J. Fateel, "Conceptual understanding, procedural knowledge and problem-solving skills in mathematics: High school graduates work analysis and standpoints," *International Journal of Education and Practice*, vol. 7, no. 3, pp. 258–273, 2019, doi: 10.18488/journal.61.2019.73.258.273.

[21] J. L. Booth and J. L. Davenport, "The role of problem representation and feature knowledge in algebraic equation-solving," *Journal of Mathematical Behavior*, vol. 32, no. 3, pp. 415–423, 2013, doi: 10.1016/j.jmathb.2013.04.003.

[22] K. H. Khng and K. Lee, "Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving," *Learning and Individual Differences*, vol. 19, no. 2, pp. 262–268, 2009, doi: 10.1016/j.lindif.2009.01.004.

[23] C. Rivella, C. Cornoldi, S. Caviola, and D. Giofrè, "Learning a new geometric concept: The role of working memory and of domain-specific abilities," *British Journal of Educational Psychology*, vol. 91, no. 4, pp. 1537–1554, 2021, doi: 10.1111/bjep.12434.

[24] T. Hailikari, A. Nevgi, and S. Lindblom-Ylänne, "Exploring alternative ways of assessing prior knowledge, Its components and their Relation to student achievement: A mathematics based case study," *Studies in Educational Evaluation*, vol. 33, no. 3–4, pp. 320–337, 2007, doi: 10.1016/j.stueduc.2007.07.007.

[25] A. K. Jitendra, A. E. Lein, J. R. Star, and D. N. Dupuis, "The contribution of domain-specific knowledge in predicting students' proportional word problem solving performance," *Educational Research and Evaluation*, vol. 19, no. 8, pp. 700-716, 2013, doi: 10.1080/13803611.2013.845107.

[26] R. G. Mohyuddin and U. Khalil, "Misconceptions of students in learning mathematics at primary level," *Bulletin of Education and Research*, vol. 38, no. 1, pp. 133–162, 2016.

[27] R. F. Jabal and R. Rosjanuardi, "Identifying the secondary school students' misconceptions about number," *Journal of Physics: Conference Series*, vol. 1157, no. 4, 2019, doi: 10.1088/1742-6596/1157/4/042052.

[28] H. R. Hj Sarwadi and M. Shahrill, "Understanding students' mathematical errors and misconceptions: The case of year 11 repeating students," *Mathematics Education Trends and Research*, vol. 2014, pp. 1–10, 2014, doi: 10.5899/2014/metr-00051.

[29] M. Oktaviani, "Analysis of students' error in doing mathematics problem on proportion," In *2nd Asian Education Symposium - AES*, 2018, pp. 172–177, doi: 10.5220/0007300601720177.

[30] G. Tan Sisman and M. Aksu, "A study on sixth grade students' misconceptions and errors in spatial measurement: Length, area, and volume," *International Journal of Science and Mathematics Education*, vol. 14, no. 7, pp. 1293–1319, 2016, doi: 10.1007/s10763-015-9642-5.

[31] P. Phonapichat, S. Wongwanich, and S. Sujiva, "An analysis of elementary school students' difficulties in mathematical problem solving," *Procedia - Social and Behavioral Sciences*, vol. 116, pp. 3169–3174, 2014, doi: 10.1016/j.sbspro.2014.01.728.

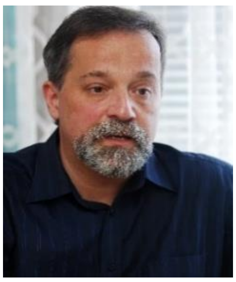
[32] K. Hyrkäs, K. Appelqvist-Schmidlechner, and L. Oksa, "Validating an instrument for clinical supervision using an expert panel," *International Journal of Nursing Studies*, vol. 40, no. 6, pp. 619–625, 2003, doi: 10.1016/S0020-7489(03)00036-1.

[33] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, 2016, doi: 10.1016/j.jcm.2016.02.012.

[34] T. G. Bond and C. M. Fox, *Applying the rasch model: Fundamental measurement in the human sciences*. New York: Routledge, 2015.

BIOGRAPHIES OF AUTHORS

Ijtihadi Kamilia Amalina is a Ph.D. candidate, Doctoral School of Education, University of Szeged, Hungary. She received her bachelor and master degrees in mathematics education from Universitas Negeri Surabaya, Indonesia. Her research focuses on assessment, mathematics education, STEM, social-emotional, and problem-solving. She can be contacted at email: ijtihadi.kamilia.amalina@edu.u-szeged.hu.



Tibor Vidákovich is a professor at the Institute of Education, University of Szeged, Hungary. He obtained his PhD in education at the Hungarian Academy of Sciences in 1991, and he was appointed a university professor in 2010. His research interests include educational assessment, diagnostic testing, development of tests and test banks; and the development and enhancement of deductive reasoning, mathematical abilities and mathematical competence. He can be contacted email: t.vidakovich@edpsy.u-szeged.hu.

Study 4: Development and differences in mathematical problem-solving skills: a cross-sectional study of differences in demographic background.

(Amalina & Vidákovich, 2023c)

Heliyon 9 (2023) e16366

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Development and differences in mathematical problem-solving skills: A cross-sectional study of differences in demographic backgrounds

Ijtihadi Kamilia Amalina [a,*], Tibor Vidákovich [b]

[a] Doctoral School of Education, University of Szeged, Hungary
[b] Institute of Education, University of Szeged, Hungary

ARTICLE INFO

Keywords:
Cross-sectional study
Demographic background
Development
Difference
Mathematical problem-solving skills

ABSTRACT

Problem-solving skills are the most applicable cognitive tool in mathematics, and improving the problem-solving skills of students is a primary aim of education. However, teachers need to know the best period of development and the differences among students to determine the best teaching and learning methods. This study aims to investigate the development and differences in mathematical problem-solving skills of students based on their grades, gender, and school locations. A scenario-based mathematical essay test was administered to 1067 students in grades 7–9 from schools in east Java, Indonesia, and their scores were converted into a logit scale for statistical analysis. The results of a one-way analysis of variance and an independent sample *t*-test showed that the students had an average level of mathematical problem-solving skills. The number of students who failed increased with the problem-solving phase. The students showed development of problem-solving skills from grade 7 to grade 8 but not in grade 9. A similar pattern of development was observed in the subsample of urban students, both male and female. The demographic background had a significant effect, as students from urban schools outperformed students from rural schools, and female students outperformed male students. The development of problem-solving skills in each phase as well as the effects of the demographic background of the participants were thoroughly examined. Further studies are needed with participants of more varied backgrounds.

1. Introduction

Problem-solving skills are a complex set of cognitive, behavioral, and attitudinal components that are situational and dependent on thorough knowledge and experience [1,2]. Problem-solving skills are acquired over time and are the most widely applicable cognitive tool [3]. Problem-solving skills are particularly important in mathematics education [3,4]. The development of mathematical problem-solving skills can differ based on age, gender stereotypes, and school locations [5–10]. Fostering the development of mathematical problem-solving skills is a major goal of educational systems because they provide a tool for success [3,11]. Mathematical problem-solving skills are developed through explicit training and enriching materials [12]. Teachers must understand how student profiles influence the development of mathematical problem-solving skills to optimize their teaching methods.

* Corresponding author. Doctoral School of Education, University of Szeged, 6722, Szeged, Petőfi S. sgt.32-34, Hungary.
E-mail addresses: ijtihadi.kamilia.amalina@edu.u-szeged.hu (I.K. Amalina), t.vidakovich@edpsy.u-szeged.hu (T. Vidákovich).

https://doi.org/10.1016/j.heliyon.2023.e16366
Received 22 December 2022; Received in revised form 5 May 2023; Accepted 14 May 2023
Available online 16 May 2023
2405-8440/

Various studies on the development of mathematical problem-solving skills have yielded mixed results. Grissom [13] concluded that problem-solving skills were fixed and immutable. Meanwhile, other researchers argued that problem-solving skills developed over time and were modifiable, providing an opportunity for their enhancement through targeted educational intervention when problem-solving skills developed quickly [3,4,12]. Tracing the development of mathematical problem-solving skills is crucial. Further, the results of previous studies are debatable, necessitating a comprehensive study in the development of students' mathematical problem-solving skills.

Differences in mathematical problem-solving skills have been identified based on gender and school location [6–10]. School location affects school segregation and school quality [9,14]. The socioeconomic and sociocultural characteristics of a residential area where a school is located are the factors affecting academic achievement [14]. Studies in several countries have shown that students in urban schools demonstrated better performance and problem-solving skills in mathematics [9,10,15]. However, contradictory results have been obtained for other countries [6,10].

Studies on gender differences have shown that male students outperform female students in mathematics, which has piqued the interest of psychologists, sociologists, and educators [7,16,17]. The differences appear to be because of brain structure; however, sociologists argue that gender equality can be achieved by providing equal educational opportunities [8,16,18,19]. Because the results are debatable and no studies on gender differences across grades in schools have been conducted, it would be interesting to investigate gender differences in mathematical problem-solving skills.

Based on the previous explanations, teachers need to understand the best time for students to develop mathematical problem-solving skills because problem-solving is an obligatory mathematics skill to be mastered. However, no relevant studies focused on Indonesia have been conducted regarding the mathematical problem-solving skill development of students in middle school that can provide the necessary information for teachers. Further, middle school is the important first phase of developing critical thinking skills; thus relevant studies are required in this case [3,4]. In addition, a municipal policy-making system can raise differences in problem-solving skills based on different demographic backgrounds [10]. Moreover, the results of previous studies regarding the development and differences in mathematical problem-solving skills are debatable. Thus, the present study has been conducted to meet these gaps. This study investigated the development of mathematical problem-solving skills in students and the differences owing demographic backgrounds. Three aspects were considered: (1) student profiles of mathematical problem-solving skills, (2) development of their mathematical problem-solving skills across grades, and (3) significant differences in mathematical problem-solving skills based on gender and school location. The results of the present study will provide detailed information regarding the subsample that contributes to the development of mathematical problem-solving skills in students based on their demographic backgrounds. In addition, the description of the score is in the form of a logit scale from large-scale data providing consistent meaning and confident generalization. This study can be used to determine appropriate teaching and learning in the best period of students' development in mathematical problem-solving skills as well as policies to achieve educational equality.

2. Theoretical background

2.1. Mathematical problem-solving skills and their development

Solving mathematical problems is a complex cognitive ability that requires students to understand the problem as well as apply mathematical concepts to them [20]. Researchers have described the phases of solving a mathematical problem as understanding the problem, devising a plan, conducting out the plan, and looking back [20–24]. Because mathematical problems are complex, students may struggle with several phases, including applying mathematical knowledge, determining the concepts to use, and stating mathematical sentences (e.g., arithmetic) [20]. Studies have concluded that more students fail at later stages of the solution process [25,26]. In other words, fewer students fail in the phase of understanding a problem than during the plan implementation phase. Different studies have stated that students face difficulties in understanding the problem, determining what to assume, and investigating relevant information [27]. This makes them unable to translate the problem into a mathematical form.

Age or grade is viewed as one factor that influences mathematical problem-solving skills because the skills of the students improve over time as a result of the teaching and learning processes [28]. Neuroscience research has shown that older students have fewer problems with arithmetic than younger students; however, the hemispheric asymmetry is reduced [29]. In other words, older students are more proficient, but their flexibility to switch among different strategies is less. Ameer & Sigh [28] obtained similar results and found a considerable difference in mathematical achievement; specifically, older students performed better than younger students in number sense and computation using one-way analysis of variance (ANOVA) (F) of $F(2,411) = 4.82$, $p < 0.01$. Molnár et al. [3] found that the student grade affects domain-specific and complex problem-solving skills. They observed that the development of problem-solving skills was noticeable across grades in elementary school but stopped in secondary school. The fastest development of domain-specific problem-solving occurred in grades 7 and 8 [3], but the fastest development of complex problem-solving occurred in grades 5–7 [3]. No development was detected between grades 4 and 5 as well as grades 6 and 7 for domain-specific and complex problem-solving skills, respectively. Similarly, Greiff et al. [4] concluded that students developed problem-solving skills across grades 5–11 with older students being more skilled. However, the grade 9 students deviated from the development pattern, and their problem-solving skills dropped. The theories from Molnár et al. [3] and Greiff et al. [4] are the benchmark cases herein.

The above studies showed that problem-solving skills mostly developed during compulsory schooling and developed most quickly in specific grades. This indicates that specific development times can be targeted to enhance the problem-solving skills [3]. However, Jabor et al. [30] observed contradictory results showing statistically significant differences with small effects in mathematical performance between age groups: those under the age of 19 outperformed those over the age of 19 years old. Grissom [13] observed a negative correlation between age and school achievement that remained constant over time.

2.2. Effects of school location and gender on mathematical problem-solving skills

School location has been shown to affect mathematical achievement [9,14]. In 15 countries, students in rural schools performed considerably worse than students in urban schools in mathematics [9,10], science and reading [9]. In addition, Nepal [15] discovered that urban students significantly outperformed rural students in mathematical problem-solving skills ($t = -5.11$, $p < 0.001$) and achievement ($t = -4.45$, $p < 0.001$) using the results of an independent sample *t*-test (t). However, other countries have found that rural students outperformed urban students in mathematics [6,10]. These variations may be attributed to a lack of instructional resources (e.g., facilities, materials, and programs), professional training (e.g., poorly trained teachers), and progressive instruction [6]. The results of Williams's study [10] serve as the basis for the current study.

Gender differences in mathematics have received attention because studies show that male students outperform female students on higher-level cognitive tasks [31]. This is a shift from a meta-analysis study that found gender differences in mathematics to be insignificant and favored female students [32]. At the college level, female students slightly outperform male students in computation while male students outperform female students in problem solving. However, no gender differences have been observed among elementary and middle school students. This result was strengthened by other meta-analysis studies [7,8], which concluded that there was no gender difference in mathematical performance and problem-solving skills [15,33–35]. Gender similarity in mathematics is achieved when equal learning opportunities and educational choices are provided and the curriculum is expanded to include the needs and interests of the students [16,18,31].

From a sociological perspective, gender similarity in mathematics makes sense. If there is a gender difference in mathematics, this has been attributed to science, technology, engineering, and mathematics (STEM) being stereotyped as a male domain [8]. Stereotypes influence beliefs and self-efficacy of students and perceptions of their own abilities [8,19]. This is the reason for the low interest of female students in advanced mathematics courses [18,19]. However, Halpern et al. [16] found that more female students are entering many occupations that require a high level of mathematical knowledge. Moreover, Anjum [36] found that female students outperformed male students in mathematics. This may be because female students prepared better than the male students before the test and were more thorough [36,37]. The study of Anjum [36] is one of the basis cases of the current study.

Differences in brain structure support the argument that there are gender differences in mathematical performance [16,17]. Females have less brain lateralization (i.e., symmetric left and right hemispheres), which helps them perform better verbally. Meanwhile, males have more brain lateralization, which is important for spatial tasks [17]. In addition, the male hormone testosterone slows the development of the left hemisphere [16], which improves the performance of right brain-dominant mathematical reasoning and spatial tasks.

3. Method

3.1. Instrumentation

In this study, a science-related mathematical problem-solving test was used. This is a mathematics essay test where the problems are in the form of scenarios related to environmental management. Problems are solved by using technology as a tool (e.g., calculator, grid paper). The test was developed in an interdisciplinary STEM framework, and it is targeted toward grades 7–9. There were six scenarios in total: some were given to multiple grades, and others were specific to a grade. They included ecofriendly packaging (grade 7), school park (grade 7), calorie vs. greenhouse gas emissions (grades 7–9), floodwater reservoir (grade 8), city park (grades 8–9), and infiltration well (grade 9). These scenarios cover topics such as number and measurement, ratio and proportion, geometry, and statistics. Every scenario had a challenge, and students were provided with eight metacognitive prompt items to help them explore their problem-solving skills.

The test was administered by using paper and pencils for a 3-h period with a break every hour. At the end of the test, students were asked to fill in their demographic information. Each prompt item had a maximum score of 5 points: a complete and correct answer (5 points), a complete answer with a minor error (4 points), an incomplete answer with a minor error (3 points), an incomplete answer with a major error (2 points), and a completely wrong and irrelevant answer (1 point). Each scenario had a maximum total score of 40 points.

The test was validated to determine whether it contained good and acceptable psychometric evidence. It had an acceptable content validity index (CVI >0.67), moderate intraclass correlation coefficient (ICC) ($r_{xx} = 0.63$), and acceptable Cronbach's alpha ($\alpha = 0.84$). The construct validity indicated all scenarios and prompt items were fit ($0.77 \leq$ weighted mean square ≤ 1.59) with an acceptable discrimination value ($0.48 \leq$ discrimination value ≤ 0.93), acceptable behavior of the rating score, and good reliability (scenario reliability = 0.86; prompt item reliability = 0.94).

3.2. Participants

The test was administered to grades 7–9 students in east Java, Indonesia (n = 1067). The students were selected from A-accreditation schools in urban and rural areas; random classes were selected for each grade. The majority of the students were Javanese (95.01%), with the remainder being Madurese (3.3%) and other ethnicities. Table 1 describes the demographics of the participants.

3.3. Data analysis

Data were collected between July and September 2022. Prior to data collection, ethical approval was sought from the institutional review board (IRB) of the Doctoral School of Education, University of Szeged and was granted with the ethical approval number of 7/2022. In addition, permission letters were sent to several schools to request permission and confirm their participation. The test answers of the students were scored by two raters – the first author of this study and a rater with master's degree in mathematics education – to ensure that the rating scale was consistently implemented. The results showed good consistency with an ICC of 0.992 and Cronbach's alpha of 0.996.

The scores from one of the raters were converted to a logit scale by weighted likelihood estimation (WLE) using the ConQuest software. A logit scale provides a consistent value or meaning in the form of intervals. The logit scale represents the unit interval between locations on the person–item map. WLE was chosen rather than maximum likelihood estimation (MLE) because WLE is more central than MLE, which helps to correct for bias [38]. The WLE scale was represented by using descriptive statistics to profile the students' mathematical problem-solving skills in terms of the percentage, mean score (*M*) and standard deviation (*SD*) for each phase. The WLE scale was also used to describe common difficulties for each phase. The development of students' mathematical problem-solving skills across grades was presented by a pirate plot, which is used in R to visualize the relationship between 1 and 3 categorical independent variables and 1 continuous dependent variable. It was chosen because it displays raw data, descriptive sta- tistics, and inferential statistics at the same time. The data analysis was performed using R studio version 4.1.3 software with the YaRrr package. A one-way ANOVA was performed to find significant differences across grades. An independent sample *t*-test was used to analyze significant differences based on gender and school location. The descriptive statistics, one-way ANOVA test, and independent sample *t*-test were performed using the IBM SPSS Statistics 25 software.

4. Results

4.1. Student profiles

The scores of students were converted to the WLE scale, where a score of zero represented a student with average ability, a positive score indicated above-average ability, and a negative score indicated below-average ability. A higher score indicated higher ability. The mean score represented a student with average mathematical problem-solving skills ($M = 0.001$, $SD = 0.39$). Overall, 52.1% of students had a score below zero. The distribution of scores among students was predominantly in the interval between —1 and 0. When the problem-solving process was analyzed by phase, the results showed that exploring and understanding were the most mastered problem-solving skills ($M = 0.24$, $SD = 0.51$). Only 27.9% of students had below-average scores for the exploring and understanding phases, which indicates that they mostly understood the given problem and recognized the important information. However, the problem-solving skills decreased with higher phases. The students had below-average abilities in the phases of representing and formulating ($M =$ —0.01, $SD = 0.36$), planning and executing ($M =$ —0.15, $SD = 0.41$), and monitoring and reflecting ($M =$ —0.16, $SD = 0.36$). About 57.9% of the students had below-average scores for the representing and formulating phase, which indicates that they had problems making hypotheses regarding science phenomena, representing problems in mathematical form, and designing a pro- totype. The obvious reason for their difficulty with making hypotheses was that they did not understand simple concepts of science (e.g., CO_2 vs. O_2). In the planning and executing phase, 66.8% of the students failed to achieve a score greater than zero. This happened because they failed to apply mathematical concepts and procedures. Because they were unable to plan and execute a strategy, this affected the next phase of the problem-solving process. In the monitoring and reflecting phase, 68.0% of the students had a below-average score.

4.2. Development of mathematical problem-solving skills across grades

The development of the mathematical problem-solving skills of the students across grades was observed based on the increase in the mean score. The problem-solving skills developed from grade 7 to grade 8. The students of grade 7 had a mean score of —0.04 while grade 8 students had the highest mean score of 0.03. The students in grades 7 and 8 also showed more varied problem-solving skills than the grade 9 students did. In contrast, the grade 9 students showed a different pattern of development, and their mean score dropped to 0.01. Although the difference was not large, further analysis was needed to determine its significance.

Fig. 1 displays the development of the mathematical problem-solving skills of the students. The dots represent raw data or WLE

Table 1
Demographic characteristics of participants.

Demographic Characteristics		N	%
Gender	Male	452	42.4
	Female	615	57.6
Grade	7 (M age = 12.59, SD = 0.61)	380	35.61
	8 (M age = 13.42, SD = 0.59)	331	31.02
	9 (M age = 14.50, SD = 0.59)	356	33.36
School location	Rural	427	40.02
	Urban	640	59.89

scores. The middle line shows the mean score. The beans represent a smoothed density curve showing the full data distribution. The scores of the students in grades 7 and 9 were concentrated in the interval between —0.5 and 0. However, the scores of the grade 8 students were concentrated in the interval between 0 and 0.5. The scores of the students in grades 7 and 8 showed a wider distribution than those of the grade 9 students. The bands which overlap with the line representing the mean score, define the inference around the mean (i.e., 95% of the data are in this interval). The inference of the WLE score was close to the mean.

The one-way ANOVA results indicated a significant difference among the problem-solving skills of the students of grades 7–9 ($F(1,066) = 3.01$, $p = 0.046$). The students of grade 8 showed a significant difference in problem-solving skills and outperformed the other students. The students of grades 7 and 9 showed no significant difference in their mathematical problem-solving skills. Table 2 presents the one-way ANOVA results of the mathematical problem-solving skills across grades.

Fig. 2 shows the development of the mathematical problem-solving skills of the students across grades based on school location and gender. The problem-solving skills of the urban students increased from a mean score of 0.07 in grade 7 to 0.14 in grade 8. However, the mean score of urban students in grade 9 dropped. In contrast, the mean scores of the rural students increased continuously with grade. The improvements were significant for both the rural ($F(426) = 10.10$, $p < 0.001$) and urban ($F(639) = 6.10$, $p < 0.01$) students. For the rural students, grade 9 students showed a significant difference in problem-solving skills. In contrast, urban students in grades 8 and 9 showed significant differences in problem-solving skills but not in grade 7.

When divided by gender, both female and male students showed improvements in their problem-solving skills from grades 7 and 8. However, female students in grade 9 showed a stable score while the scores of male students in grade 9 declined. Only male students in grade 7 showed a significant difference in the mean score. In urban schools, the scores of male and female students increased and decreased, respectively, from grade 7 to grade 8. Male students in rural schools showed an increase in score from grade 7 to grade 9. However, the scores of female students in rural schools decreased from grade 7 to grade 8. Table 3 presents the one-way ANOVA results for the mathematical problem-solving skills of the students considering gender and school location.

Fig. 2 shows that the distributions of the male and female scores of students were similar for every grade except rural grade 9 students. The scores of the rural female students were concentrated in the interval between 0 and 0.5 while the scores of the rural male students were mostly below 0. The scores of rural students in grade 7 and urban students in grade 9 (both male and female) were concentrated in the interval between —0.5 and 0. The scores of urban students in grades 7 and 8 were concentrated in the interval between —0.5 and 0.5.

Fig. 3 shows a detailed analysis of the development of mathematical problem-solving skills across grades for each phase of the problem-solving process. Similar patterns were observed in the exploring and understanding and the representing and formulating phases: the mean score increased from grade 7 to grade 8 but decreased from grade 8 to grade 9. Grade 8 students had the highest mean score and differed significantly from the scores of students in other grades.

The scores of the students for the planning and executing phase increased with grade. However, the difference was only significant at grade 9. Grades 7 and 8 students showed an increase in score, but the improvement was not significant. There was no pattern detected in the monitoring and reflecting phase. The score was stable for grades 7 and 8 students but improved for grade 9 students. The mean score for each phase and the one-way ANOVA results are presented in Table 4.

Fig. 3 shows that the distributions of the problem-solving skills of the students were similar across grades and phases. However, the distributions were different for grade 9 students in the representing and formulating, planning and executing, and monitoring and reflecting phases, where 95% of the data were in the interval between —0.5 and 0.5.

4.3. Effects of demographic background

4.3.1. School location

The mathematical problem-solving skills of the students differed significantly based on school location. Urban students scored higher than rural students. The results of the *t*-test for mathematical problem-solving skills based on school location are presented in Table 5.

The effects of the school's location on the performances of male and female students were analyzed. The results showed that the

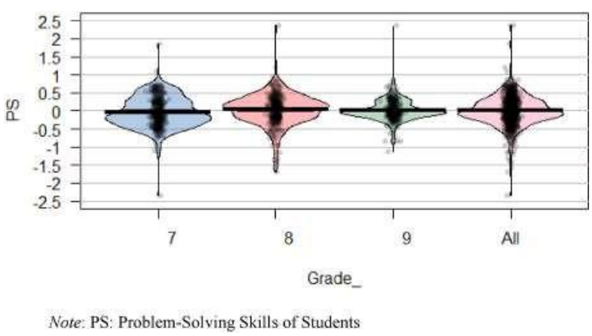


Fig. 1. Differences in students' mathematical problem-solving skills across grades.
Note: PS: Problem-Solving Skills of Students.

Table 2
One-way ANOVA results of the mathematical problem-solving across grades.

Grades	N	*M*	*SD*	*F*(1,066)	*P*	Significant difference between the sub-samples
7	380	−0.04	0.42	3.01	.046	{7, 9} < {8}
8	331	0.03	0.44			
9	356	0.01	0.29			

Note. Post hoc test: Dunnett's T3. 7, 8, and 9: subsample grade. <: direction of significant difference ($p < 0.05$).

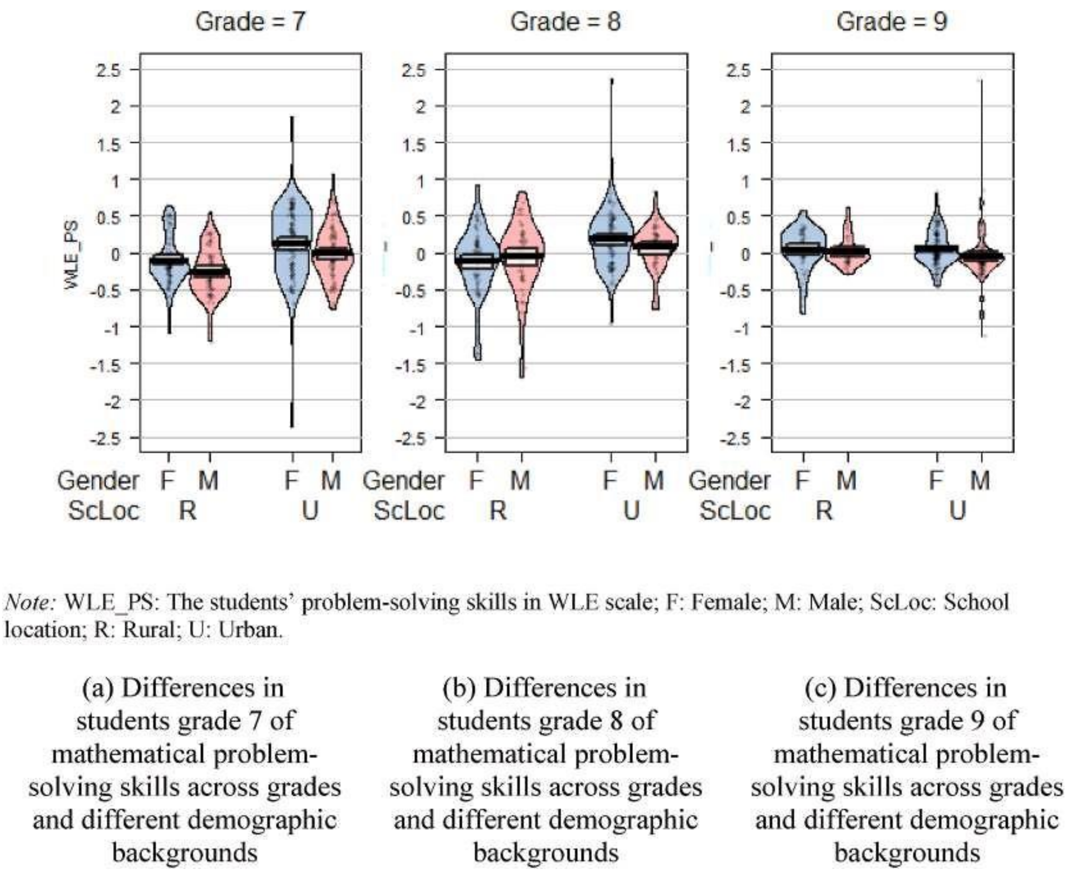


Fig. 2. Differences in students' mathematical problem-solving skills across grades and different demographic backgrounds.
(a) Differences in students grade 7 of mathematical problem-solving skills across grades and different demographic backgrounds
(b) Differences in students grade 8 of mathematical problem-solving skills across grades and different demographic backgrounds
(c) Differences in students grade 9 of mathematical problem-solving skills across grades and different demographic backgrounds
Note: WLE_PS: The students' problem-solving skills in WLE scale; F: Female; M: Male; ScLoc: School location; R: Rural; U: Urban.

scores of the female students differed significantly based on school location ($t(613) = -6.09$, $p < 0.001$). Female students in urban schools ($M = 0.18$, $SD = 0.39$) outperformed female students in rural schools ($M = -0.08$, $SD = 0.37$). Similar results were observed for male students with urban students ($M = -0.01$, $SD = 0.35$) outperforming rural students ($M = -0.12$, $SD = 0.39$) by a significant margin ($t(382.764) = -3.25$, $p < 0.01$).

When analyzed by grade, grades 7 and 8 students contributed to the difference based on school location with $t(377.952) = -6.34$, $p < 0.001$ and $t(300.070) = -5.04$, $p < 0.001$, respectively. Urban students in grades 7 and 8 performed significantly better than their rural counterparts did. However, there was no significant difference between rural and urban students in grade 9 ($t(354) = 0.71$, $p = 0.447$).

4.3.2. Gender

Male and female students showed a significant difference in their mathematical problem-solving skills. Overall, female students outperformed male students. The detailed results of the independent sample *t*-test for mathematical problem-solving skills based on gender are presented in Table 6.

The results were analyzed to determine whether the school location contributed to the gender difference. The gender difference was

Table 3
One-way ANOVA results for mathematical problem-solving skills across grades and different demographic backgrounds.

	Grade	*N*	*M*	*SD*	*F*	*P*	Significant difference between the sub-samples
Rural	7	168	−0.18	0.34	*F(426) = 10.10*	*<.001*	{7, 8} < {9}
	8	155	−0.10	0.46			
	9	104	0.03	0.25			
Urban	7	212	0.07	0.44	F(639) = 6.10	.001	{7, 9} < {8}
	8	176	0.14	0.38			
	9	252	0.05	0.30			
Female	7	219	0.03	0.44	F(614) = 0.18	.84	–
	8	204	0.05	0.45			
	9	192	0.05	0.26			
Male	7	161	−0.13	0.36	F(451) = 5.28	.005	{7} < {8, 9}
	8	127	0.001	0.42			
	9	164	−0.03	0.32			

Note. Post hoc test: Dunnett's T3. 7, 8, and 9: subsample grade. <: direction of significant difference ($p < 0.05$).

most significant among urban students ($t(596.796) = —4.36$, $p < 0.001$). Female students from urban schools ($M = 0.12$, $SD = 0.39$) outperformed male students from urban schools ($M = —0.01$, $SD = 0.35$). There was no significant difference between female and male students from rural schools ($t(425) = —1.31$, $p = 0.191$).

Grades 7 and 9 students contributed to the gender difference with $t(372.996) = —3.90$, $p < 0.001$ and $t(354) = —2.73$, $p < 0.01$, respectively. Female students in grades 7 and 9 outperformed their male counterparts. However, there was no significant gender difference among grade 8 students ($t(329) = —0.10$, $p = 0.323$).

5. Discussion

The mathematical problem-solving skills of the students were categorized as average. In addition, the difficulties of students increased in line with the problem-solving phase. Fewer students failed the exploring and understanding phase than the subsequent phases. This confirms the results of previous studies indicating that more students failed further along the problem-solving process [25,26]. Because the problem-solving process is sequential, students who have difficulty understanding a problem will fail the subsequent phases [27].

The development of mathematical problem-solving skills was evaluated according to the mean WLE score. The mathematical problem-solving skills of the students developed from grade 7 to grade 8 based on the increase in their mean scores. However, the development dropped in grade 9. This agrees with previous results that concluded that higher grades had the highest problem-solving skills, but the fastest skill development took place in grades 7–8 after which it dropped [3,4]. These results indicate that the mathematical problem-solving skills of the students should improve and be strengthened in grades 7–8, which will help them perform better in grade 9.

In this study, the effects of the demographic background of the students were analyzed in detail, which is an aspect missing from previous studies. The results showed that the mathematical problem-solving skills of urban students increased from grade 7 to grade 8 but decreased in grade 9. The same pattern was found among male and female students. However, a different pattern was observed for rural students, where the skills of grade 9 students continued to increase. The different patterns may be attributed to a structural reorganization of cognitive processes at a particular age [3]. However, more research is needed on the effects of the demographic backgrounds of students on mathematical problem-solving skills. These results were different from previous results because the previous studies only analyzed the development in general, without focusing on their demographic background. Hence, different patterns of development were observed when it was thoroughly examined.

Because solving problems is a cognitive process, the development of problem-solving skills for particular phases and processes needed to be analyzed. The students showed the same pattern for knowledge acquisition (i.e., exploring and understanding, and representing and formulating phases), with an increase in skill from grade 7 to grade 8 but a decrease in grade 9. However, the students showed increasing skill in knowledge application (i.e., planning and executing, as well as monitoring and reflecting phases) across grades. This means that the difference between the mean scores in grade 9 was not significant across phases. Grade 9 students had lower scores than grade 8 students for the knowledge acquisition phase but higher scores for the knowledge application phase. In contrast, the gap between the mean scores of grades 7 and 8 was large across phases.

These results proved that there is a significant difference in the mathematical problem-solving skills of students based on their demographic backgrounds. The urban students outperformed rural students, which confirms the results of previous studies [9,10,15]. The difference can be attributed to the availability of facilities, teacher quality, and interactive teaching and learning instruction [6]. In Indonesia, the policies for the public educational system for middle schools are set at the municipal level. This means that each city has its own policies for teacher training, teacher recruitment, teaching and learning processes, facilities, etc. Urban schools mostly have stricter policies as well as various programs to help students improve their knowledge and skills. In addition, they have supportive facilities for teaching and learning. This unequal environment is the strongest reason for the difference in mathematical problem-solving skills.

The results were analyzed in detail to observe which groups in the rural and urban schools contributed to the difference. Both male and female students in urban schools performed better than their counterparts in rural schools did. In addition, urban students in

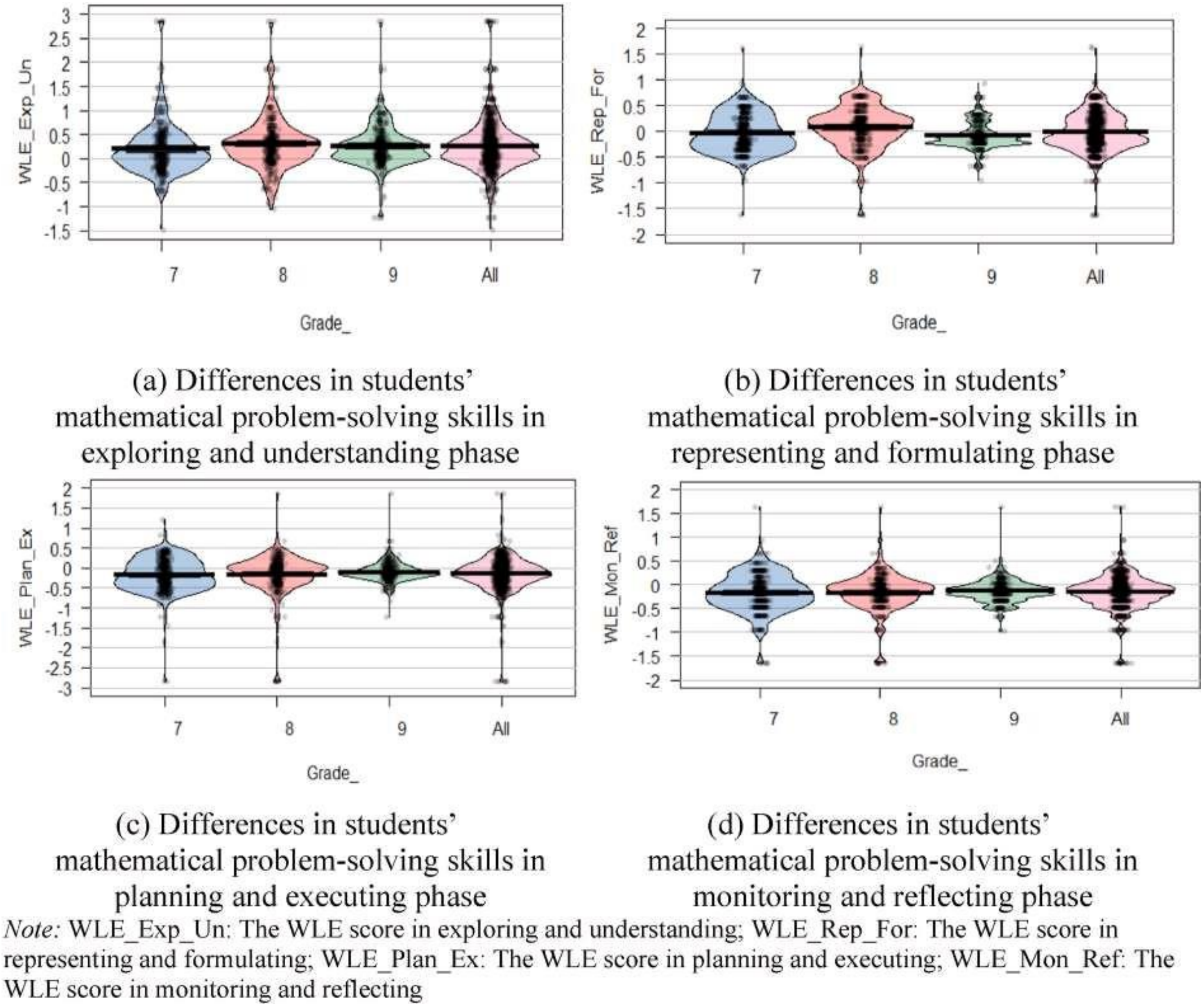


Fig. 3. Differences in students' mathematical problem-solving skills in every phase across grades: (1) Exploring & understanding, (2) Representing & formulating, (3) Planning & executing, (4) Monitoring & reflecting.
(a) Differences in students' mathematical problem-solving skills in exploring and understanding phase
(b) Differences in students' mathematical problem-solving skills in representing and formulating phase
(c) Differences in students' mathematical problem-solving skills in planning and executing phase
(d) Differences in students' mathematical problem-solving skills in monitoring and reflecting phase
Note: WLE_Exp_Un: The WLE score in exploring and understanding; WLE_Rep_For: The WLE score in representing and formulating; WLE_Plan_Ex: The WLE score in planning and executing; WLE_Mon_Ref: The WLE score in monitoring and reflecting.

grades 7 and 8 outperformed their rural counterparts. There was no significant difference between urban and rural students in grade 9. This may be because grade 9 is the last grade in middle school, so students have to prepare for high school entrance requirements, including exam and/or grade point average scores. Hence, both rural and urban schools focus much effort on the teaching and learning process in this grade.

In this study, the female students surprisingly had better mathematical problem-solving skills than the male students did. This confirmed the results of the meta-analysis by Hyde et al. [32] and study by Anjum [36], which found that female students slightly outperformed male students in mathematics. This difference may be because of motivation and attitude [39,40]. Female Indonesian students are typically more diligent, thorough, responsible, persistent, and serious with their tasks.

A detailed analysis was performed to evaluate which group of students contributed to the gender differences. The results showed that female students outperformed male students in urban schools. This may be because male students at urban schools typically display an unserious attitude toward low-stake tests. In addition, female students outperformed their male counterparts in grades 7 and 9. The reason for this difference requires further analysis.

Table 4
One-way ANOVA results for every phase of problem-solving across grades.

Phases	Grade	*M*	*SD*	*F*(1,066)	*P*	
Exploring & Understanding	7	0.19	0.54	3.98	.019	{7, 9} < {8}
	8	0.30	0.54			
	9	0.25	0.44			
Representing & Formulating	7	−0.03	0.37	16.20	<.001	{7, 9} < {8}
	8	0.07	0.42			
	9	−0.08	0.26			
Planning & Executing	7	−0.18	0.42	3.20	.041	{9} < {7, 8}
	8	−0.17	0.51			
	9	−0.11	0.24			
Monitoring & Reflecting	7	−0.17	0.41	1.48	.228	–
	8	−0.17	0.41			
	9	−0.13	0.23			

Note. Post hoc test: Dunnett's T3. 7, 8, and 9: subsample grade. <: direction of significant difference ($p < 0.05$).

Table 5
T-test results for mathematical problem-solving skills based on school location.

Location	*N*	*M*	*SD*	*F*	*P*	*t*(1065)	*P*
Rural	427	−0.10	0.38	0.18	.674	−6.90	<.001
Urban	640	0.07	0.38				

Table 6
T-test results for mathematical problem-solving skills based on gender.

Gender	*N*	*M*	*SD*	*F*	*P*	*t*(1006,013)	*P*
Female	615	0.04	0.39	5.20	.023	−4.312	<.001
Male	452	−0.06	0.38				

6. Conclusion

Studying the problem-solving skills of students is crucial to facilitating their development. In this study, the conclusions are presented as follows:

- The mathematical problem-solving skills of the students were categorized as average. More students failed at higher phases of the problem-solving process.
- Students showed development of their mathematical problem-solving skills from grade 7 to grade 8 but a decline in grade 9. The same pattern was detected across grades for urban students, both female and male. However, the problem-solving skills of rural students increased with the grade.
- A similar development was observed for the individual problem-solving phases. In the knowledge acquisition phase, the problem-solving skills of the students developed from grade 7 to grade 8 but decreased in grade 9. However, problem-solving skills increased across grades in the knowledge application phase.
- The school location was shown to have a significant effect on the mathematical problem-solving skills of the students. Urban students generally outperform students in rural schools. However, gender and grade contributed to differences in mathematical problem-solving skills based on school location. Female and male urban students in grades 7 and 8 outperformed their rural counterparts.
- In general, female students outperformed male students in mathematical problem-solving skills, particularly those from urban schools and in grades 7 and 9.

The sampling method and the number of demographic backgrounds limited the scope of this study. Only students from A-accreditation schools were selected because higher-order problem-solving skills were considered assets. Moreover, the study only included three demographic factors: grade, gender, and school location. More demographic information, such as school type, can be added (public or private schools). Hence, future studies will need to broaden the sample size and consider more demographic factors. Despite these limitations, this study can help teachers determine the best period for enhancing the development of mathematical problem-solving skills. Moreover, the differences in mathematical problem-solving skills due to demographic background can be used as a basis for educational policymakers and teachers to provide equal opportunity and equitable education to students.

Author contribution statement

Ijtihadi Kamilia Amalina: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Tibor Vidákovich: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by University of Szeged Open Access Fund with the grant number of 6020.

Data availability statement

Data will be made available on request.

Additional information

No additional information is available for this paper.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

References

[1] J. Josep, V.S. Lo´pez, Cognitive variables in science problem solving: a review of research, J. Physic Tchr. Educ. Online 4 (2) (2007) 25–32 [Online]. Available: https://roderic.uv.es/bitstream/handle/10550/43424/038260.pdf;sequence=1.
[2] B. Rittle-Johnson, J.R. Star, K. Durkin, The importance of prior knowledge when comparing examples: influences on conceptual and procedural knowledge of equation solving, J. Educ. Psychol. 101 (4) (2009) 836–852, https://doi.org/10.1037/a0016026.
[3] G. Moln´ar, S. Greiff, B. Csapo´, Inductive reasoning, domain specific and complex problem solving: relations and development, Think. Skills Creativ. 9 (2013) 35–45, https://doi.org/10.1016/j.tsc.2013.03.002.
[4] S. Greiff, S. Wüstenberg, G. Moln´ar, A. Fischer, J. Funke, B. Csapo´, Complex problem solving in educational contexts-something beyond g: concept, assessment, measurement invariance, and construct validity, J. Educ. Psychol. 105 (2) (2013) 364–379, https://doi.org/10.1037/a0031856.
[5] L. Ayebale, G. Habaasa, S. Tweheyo, Factors affecting students' achievement in mathematics in secondary schools in developing countries: a rapid systematic, Stat. J. IAOS 36 (S1) (2020) S73–S76, https://doi.org/10.3233/sji-200713.
[6] J. Lee, W.G. McIntire, Interstate variation in the mathematics achievement of rural and nonrural students, J. Res. Rural Educ. 16 (3) (2000) 168–181 [Online]. Available: https://jrre.psu.edu/sites/default/files/2019-08/16-3_2.pdf.
[7] M. Li, Y. Zhang, H. Liu, Y. Hao, Gender differences in mathematics achievement in Beijing: a meta-analysis, Br. J. Educ. Psychol. 88 (4) (2018) 566–583, https://doi.org/10.1111/bjep.12203.
[8] S.M. Lindberg, J.S. Hyde, J.L. Petersen, M.C. Linn, New trends in gender and mathematics performance: a meta analysis, Psychol. Bull. 136 (6) (2010) 1123–1135, https://doi.org/10.1037/a0021276. New.
[9] R. Ramos, J.C.C. Duque, S. Nieto, Decomposing the rural-urban differential in student achievement in Colombia using PISA microdata, SSRN Electron. J. (2021), https://doi.org/10.2139/ssrn.2051358.
[10] J.H. Williams, Cross-national variations in rural mathematics achievement: a descriptive overview, J. Res. Rural Educ. 20 (5) (2005) 1–18 [Online]. Available: https://jrre.psu.edu/sites/default/files/2019-08/20-5.pdf.
[11] OECD, *PISA*, Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, OECD Publishi, 2012 ng, 2013.
[12] P. Adey, B. Csapo´, A. Demetriou, J. Hautama¨ki, M. Shayer, Can we be intelligent about intelligence?. Why education needs the concept of plastic general ability, Educ. Res. Rev. 2 (2) (2007) 75–97, https://doi.org/10.1016/j.edurev.2007.05.001.
[13] J.B. Grissom, Age and achievement, Educ. Pol. Anal. Arch. 12 (49) (2004), https://doi.org/10.14507/epaa.v12n49, 2004.
[14] E. Tomul, E. O¨nder, E. Taslidere, The relative effect of student, family and school-related factors on math achievement by location of the school, Large-scale Assessments Educ. 9 (1) (2021) 1–29, https://doi.org/10.1186/s40536-021-00117-1.
[15] B. Nepal, Impact of gender and location on mathematical thinking and mathematics achievement, J. Adv. Acad. Res. 3 (3) (2017) 11–21, https://doi.org/10.3126/jaar.v3i3.16803.
[16] D.F. Halpern, J. Wai, A. Saw, A Psychobiosocial Model: Why Females Are Sometimes Greater than Adn Sometimes Less than Males in Math Achievement, Cambridge University Press, 2005.
[17] Z. Zhu, Gender differences in mathematical problem solving patterns: a review of literature, Int. Educ. J. 8 (2) (2007) 187–203 [Online]. Available: https://files.eric.ed.gov/fulltext/EJ834219.pdf.
[18] S. Catsambis, The gender gap in mathematics: merely a step function? in: A.M. Gallagher, J.C. Kaufman (Eds.), Gender Differences in Mathematics Cambridge University Press, 2005.
[19] P.G. Davies, S.J. Spencer, The gender-gap artifact: women's underperformance in quantitative domains through the lens of stereotype treat, in: A.M. Gallagher, J.C. Kaufman (Eds.), Gender Differences in Mathematics, Cambridge University Press, 2005.
[20] T. Tambychik, T.S.M. Meerah, Students' difficulties in mathematics problem-solving: what do they say? Procedia - Soc. Behav. Sci. 8 (August) (2010) 142–151, https://doi.org/10.1016/j.sbspro.2010.12.020.
[21] G. Polya, How to Solve it: a New Aspect of Mathematical Method, 30, second ed., The Mathematical Gazette, 1978, p. 181 [Online]. Available: http://www.jstor.org/stable/3609122?origin=crossref.
[22] I. Bayazit, An investigation of problem solving approaches, strategies, and models used by the 7th and 8th grade students when solving real-world problems, Kuram ve Uygulamada Egit. Bilim. 13 (3) (2013) 1920–1927, https://doi.org/10.12738/estp.2013.3.1419.
[23] D.A. Stylianou, An examination of middle school students' representation practices in mathematical problem solving through the lens of expert work: towards an organizing scheme, Educ. Stud. Math. 76 (3) (2011) 265–280, https://doi.org/10.1007/s10649-010-9273-2.
[24] M. Wu, R.J. Adams, Modelling mathematics problem solving item responses using a multidimensional IRT model, Math. Educ. Res. J. 18 (2) (2006) 93–113, https://doi.org/10.1007/BF03217438.
[25] M.T. Siniguian, Students difficulty in solving mathematical problems, Int. J. Adv. Res. Eng. Appl. Sci. 6 (2) (2017) 1–12 [Online]. Available: https://garph.co.uk/IJAREAS/Feb2017/1.pdf.

[26] M.R. Novriani, E. Surya, Analysis of student difficulties in mathematics problem solving ability at MTs SWASTA IRA medan, Int. J. Sci. Basic Appl. Res. 33 (3) (2017) 63–75 [Online]. Available: https://core.ac.uk/download/pdf/249335829.pdf.
[27] P. Phonapichat, S. Wongwanich, S. Sujiva, An analysis of elementary school students' difficulties in mathematical problem solving, February, Procedia - Soc. Behav. Sci. 116 (2014) 3169–3174, https://doi.org/10.1016/j.sbspro.2014.01.728.
[28] I.S. Ameer, P. Singh, Exploring Grade Levels and Gender Differences in Numeracy Thinking Among Secondary School Students," Procedia - Soc. Behav. Sci vol. 90, InCULT 2012, 2013, pp. 187–195, https://doi.org/10.1016/j.sbspro.2013.07.081.
[29] R. El Yagoubi, P. Lemaire, M. Besson, Effects of aging on arithmetic problem-solving: an event-related brain potential study, J. Cognit. Neurosci. 17 (1) (2005) 37–50, https://doi.org/10.1162/0898929052880084.
[30] M.K. Jabor, K. Machtmes, K. Kungu, Y. Buntat, M.S. Nordin, The influence of age and gender on the students ' achievement in mathematics, 2011 Int. Conf. Soc. Sci. Humanit. 5 (2011) 304–308.
[31] G.C. Leder, Gender and Mathematics Education: Overview, Springer Open, 2019.
[32] J.S. Hyde, E. Fennema, S.J. Lamon, Gender differences in mathematics performance: a meta-analysis, Psychol. Bull. 107 (2) (1990) 139–155 [Online]. Available: https://psycnet.apa.org/doi/10.1037/0033-2909.107.2.139.
[33] J.T. Ajai, B.I. Imoko, Gender differences in mathematics achievement and retention scores: a case of problem-based learning method, Int. J. Res. Educ. Sci. 1 (1) (2015) 45–50, https://doi.org/10.21890/ijres.76785.
[34] L. Salihu, M. Aro, P. R¨as¨anen, Children with learning difficulties in mathematics: relating mathematics skills and reading comprehension, Issues Educ. Res. 28 (4) (2018) 1024–1038 [Online]. Available: https://www.semanticscholar.org/paper/Children-with-learning-difficulties-in-mathematics%3A-Salihu-Aro/5a9c7f13502ebc431d07f769331763030cfdcd4e.
[35] P.M. Vilenius-Tuohimaa, K. Aunola, J.E. Nurmi, The association between mathematical word problems and reading comprehension, Educ. Psychol. 28 (4) (2008) 409–426, https://doi.org/10.1080/01443410701708228.
[36] S. Anjum, Gender difference in mathematics achievement and its relation with reading comprehension of children at upper primary stage, J. Educ. Pract. 6 (16) (2015) 71–75 [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1079951&site=ehost-live.
[37] C. Good, J. Aronson, J.A. Harder, Problems in the pipeline: stereotype threat and women's achievement in high-level math courses, J. Appl. Dev. Psychol. 29 (1) (2008) 17–28, https://doi.org/10.1016/j.appdev.2007.10.004.
[38] S. Wang, T. Wang, Precision of warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing, Appl. Psychol. Meas. 25 (4) (2001) 317–331, https://doi.org/10.1177/01466210122032163.
[39] B. Guven, B.O. Cabakcor, Factors influencing mathematical problem-solving achievement of seventh grade Turkish students, Learn. Indiv Differ 23 (1) (2013) 131–137, https://doi.org/10.1016/j.lindif.2012.10.003.
[40] M. Nicolaidou, G. Philippou, Attitudes towards mathematics, self-efficacy and achievement in problem-solving, Eur. Res. Math. Educ. III (2002) 1–11 [Online]. Available: http://www.mathematik.tu-dortmund.de/~erme/CERME3/Groups/TG2/TG2_nicolaidou_cerme3.pdf.

Study 5: Cognitive and socioeconomic factors that influence the mathematical problem-solving skills of students.

(Amalina & Vidákovich, 2023b)

Heliyon 9 (2023) e19539

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Cognitive and socioeconomic factors that influence the mathematical problem-solving skills of students

Ijtihadi Kamilia Amalina [a,*], Tibor Vidákovich [b]

[a] Doctoral School of Education, University of Szeged, 6722 Szeged, Peto˝fi S. sgt.32-34, Hungary
[b] Institute of Education, University of Szeged, Hungary

A R T I C L E I N F O

Keywords:
External factor
Internal factor
Mathematics thinking skills
Middle-school students
Structural equation modeling

A B S T R A C T

Mathematical problem-solving is necessary to encounter professional, 21st-century, and everyday challenges. The relevant context of mathematical problem-solving is related to science, which is presented using natural language. Mathematical problem-solving requires both mathematical skills and nonmathematical skills, e.g., science knowledge and text comprehension skills. Thus, several internal and external factors affect success in mathematical problem-solving. In this study, we investigated the cognitive (i.e., mathematics domain-specific prior knowledge (DSPK), science background knowledge, and text comprehension skills) and socioeconomic status (SES) (i.e., parents' educational level and family income) factors that affect students' mathematical problem-solving skills. The data considered in this study included tests, documents, and a questionnaire from grade seven to nine students (n = 1067). In addition, a theoretical model was constructed using structural equation modeling. We found that this model was close to satisfying the critical values of fit indices. The model was then modified by deleting the nonsignificant paths, and the modified model exhibited a better fit. We found that most of the exploratory variables directly affected mathematical problem-solving skills, with the exception of the parents' educational levels. The strongest factor was mathematics DSPK. Both the father's and mother's educational levels indirectly influenced mathematical problem-solving skills through family income. In addition, text comprehension skills indirectly impacted mathematical problem-solving skills with science background knowledge acting as a mediator.

1. Introduction

Education is the foundation of development and the essential tool for modern society [1,2]. Mathematics is a crucial skill that allows humans to understand the environment and obtain accurate accounts of physical phenomena [1,3]. Thus, mathematics education is critically important for lifespan [1,2].

The dynamic changing environment requires students to possess contemporary and transferable skills [4]. In addition, it is important to infuse mathematics into different subjects, such as science, because mathematics is a fundamental tool for understanding science, engineering, technology, and economics [1,4,5]. The science, technology, engineering, and mathematics (STEM) framework facilitates multidisciplinary education to solve a dynamic challenging environment [6]. Mathematics and science play a bidirectional and pivotal role; the former provides tools that can be used to represent, model, calculate, and predict quantitative relationships in

* Corresponding author.
E-mail address: ijtihadi.kamilia.amalina@edu.u-szeged.hu (I.K. Amalina).

https://doi.org/10.1016/j.heliyon.2023.e19539
Received 15 December 2022; Received in revised form 19 August 2023; Accepted 25 August 2023
Available online 29 August 2023

science, while the latter provides a rich context in mathematics [5].

General reasoning ability, inductive and deductive reasoning, and divergent and convergent thinking are essential in mathematics that are assessed through mathematical problem-solving [7]. Problem-solving is a mandatory mathematics skill and the most suitable cognitive tool in mathematics education relevant to encounter professional, twenty-first-century, and everyday challenges [6,8,9]. Problem-solving is a cognitive or mental activity (including content and process) focused on transforming any situation into a goal situation when there are no direct ways to reach the goal [10,11]. Mathematical problem-solving is a thinking process wherein a solver tries to understand the problem situation using mathematical knowledge to obtain new information about the situation in order to reach the goal [11]. Hence, previous experience and prior mathematical knowledge are required to identify a problem [2]. Mathematical problem-solving has two main characteristics, namely, the task and process [11]. The task in mathematical problem-solving involves using realistic context that is presented through natural language in the form of a mathematics word problem [12]. The relevant realistic context of mathematical problem-solving is related to science [5]. In addition, there are overlapping frameworks among general problem-solving, science, and mathematics frameworks [6,9]. However, mathematical problem-solving in the science context has been identified by students as the most difficult to understand [13]. Thus, assessing students' mathematical problem-solving skills using a science context is applicable and necessary for contemporary education.

Students struggle with mathematical problem-solving even if they have mastered the mathematical computation skills required to solve the problem because they must identify both the relevant numerical information and the corresponding linguistic information in the text [12,14,15]. In other words, factors other than mathematical knowledge and skills influence mathematical problem-solving skills. Several researchers have discovered factors influencing mathematical problem-solving, namely, internal factors, such as resource, heuristic and method, control, and affect (belief, emotion, attitude, etc.), as well as external factors, such as demographic status, socioeconomic status (SES), and teachers' expectations [16–19]. Resource refers to formal and informal conceptual and procedural understanding and it is related to cognitive factors, including domain-specific prior knowledge (DSPK), text or reading comprehension, verbal ability, quantitative ability, etc. [16]. However, a study investigated the factors influencing mathematical problem-solving, including students' characteristics, family characteristics, parents' involvement, and school characteristics [20]. The most influential factors were family and students' characteristics, which accounted for 29.9% and 18% of variances, respectively. Family characteristics included family SES (parents' educational level, family income, etc.), and student characteristics included students' knowledge (cognitive). Hence, this study focuses on cognitive and SES factors influencing mathematical problem-solving skills.

The cognitive factors that influence mathematical problem-solving depend on the problem's structure and content [21]. Thus, students will succeed in mathematics problem-solving in the science context due to their prior knowledge and experience in mathematics and their nonmathematical knowledge, such as science knowledge [6,9] and text comprehension [7,12,14,22]. In addition to cognitive factors, external factors are also correlated with students' mathematical problem-solving skills, e.g., SES factors [2,23,24]. Parents' educational level, family income, and parents' occupations are indicators of SES [1,25–27].

Studies have proven that there is a strong correlation ($r \geq 0.4$) between mathematics knowledge and science knowledge in problem-solving because domain-specific knowledge is required in the problem-solving process [6,9,19]. In addition, several studies have demonstrated the importance of text comprehension in mathematical problem-solving [7,12,15] and science achievement [28–30]. However, these studies generally failed to consider all independent variables together using structural-equation modeling (SEM). In addition, many studies have confirmed the critical effect of the parents' educational level and family income as SES indicators on mathematical problem-solving [2,23,24]. However, they considered these variables as unconnected even though they are strongly related, e.g., the parents' educational level can affect family income [25,26].

Thus, this paper examines the cognitive and SES factors that influence science-related mathematical problem-solving skills. The cognitive factor focuses on mathematics DSPK, science background knowledge, and text comprehension skills. The SES factor emphasizes the parents' educational level and family income. The cognitive and SES variables will be directly and/or indirectly correlated based on previous studies, called a theoretical model. Hence, this study proposes the following research question. "To what extent do mathematics DSPK, science knowledge, text comprehension skills, parents' educational level, and family income variables interact in predicting students' mathematical problem-solving skills?"

The current study identifies these factors in a model analyzed by SEM, which can measure the simultaneous testing of relationships (both direct and indirect) rather than a single separate factor. The new model generated in this study is formulated based on previous studies. In addition, we compare the results from several developed models for more accurate and convincing results.

2. Theoretical background

A problem has been described as a gap between a current situation and a goal state [31]. Problem-solving is a cognitive process to understand, transform, and solve a problem when the solution is not immediately obvious [6,8]. Problem-solving is considered to be a process rather than the result itself [6].

Solving a problem represented using natural language by identifying important information in the text and applying mathematical knowledge is referred to as solving a mathematical word problem [14,32]. There are five foundational skills in mathematics education, that is, conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition [14]. When solving a mathematical word problem, most of these foundational skills are employed, such as, describing a real-world situation (productive disposition), applying concepts and procedures, using an appropriate strategy, and performing logical reasoning (strategic reasoning).

2.1. Mathematics DSPK and science background knowledge in mathematical problem-solving skills

Mathematics and science knowledge is highly related to reasoning and problem-solving [6], and the frameworks for general problem-solving, mathematical problem-solving, and science problem-solving overlap [6,9]. The overlapping of cognitive processes in mathematics and problem-solving involve structuring, mathematizing, processing, interpreting, and validating [5,9]. In addition, science is also related to problem-solving in terms of the application of knowledge, the performance of the inquiry process, and the relevant given context [9]. Thus, to realize successful mathematical problem-solving using the science context, students must master both mathematics knowledge and science knowledge.

The important role mathematics and science knowledge plays in problem-solving has been confirmed by several studies [6,9]. For example, a previous study demonstrated that mathematics and science are strongly correlated with complex problem-solving, exhibiting correlation coefficients of .492 and .401, respectively [6]. In addition, a study covering 41 countries investigated the contribution of mathematics and science competencies in analytical problem-solving and found that these competencies significantly contribute to problem-solving [9]. The authors also introduced the “math-science coherence” concept, which refers to the degree to which science and mathematics are harmonized as a feature of the educational environment.

Students’ prior mathematics knowledge is significantly related to mathematical problem-solving [19,33]. Prior mathematical knowledge includes knowledge related to concepts, facts, definitions, relations, and operations (i.e., conceptual knowledge) and knowledge related to how and when to use the relevant procedures (i.e., procedural knowledge) [34,35].

Previous studies have revealed that conceptual and procedural knowledge influence mathematical problem-solving [35–38]. Mathematics DSPK accounted for 56% of the variance of problem-solving [37]. A study into algebraic problem-solving revealed that conceptual understanding in algebra explained 31% of the variance of problem-solving score [35]. A similar study concluded that conceptual knowledge about fractions was the most correlated factor ($r = 0.45$) influencing proportion problem-solving [36]. Moreover, another study reported that the ratio of students’ mastery of prior algebra knowledge significantly predicted their achievement in solving algebra problem [19].

2.2. Text comprehension skills in mathematical problem-solving skills

Solving mathematical word problems requires the solver to identify the relevant linguistic and numerical information contained in the text [14,32]. Thus, it is necessary to possess both mathematical and linguistic skills (e.g., text comprehension skills) because students must comprehend the language of the text before applying an appropriate algorithm [14,22]. Students use their knowledge structures and networks by linking their prior knowledge to the information and concepts conveyed in a text [22].

Text comprehension is the construction of mental representation of the textual information to comprehend the text and merge it with prior knowledge [3,9,14]. Text comprehension skills include understanding the vocabulary of a context, identifying the main concept, noting detail, making inference, predicting outcomes, and forming conclusions [13]. Mathematics text differs from narrative or expository because mathematics employs a unique language, symbols, and vocabulary [3,22].

Studies have proven that text comprehension skills are strongly correlated to mathematical problem-solving [12,15,39]. For example, the correlation coefficients between mathematical word problem-solving skill and text comprehension skill were 0.45 and 0.38, respectively [15,39]. However, the model in a study involved a symmetric effect between mathematical problem-solving skills and text comprehension skills [39]. Mathematics problem-solving also affects mathematics achievement [7]. Text comprehension skills are correlated with mathematics achievement [3,14,40,41]. The text comprehension skills and mathematics achievement are demonstrating a correlation coefficient of 0.473 [3].

2.3. Text comprehension skills in science achievement

Text comprehension skills and science skills are interrelated, for example, identifying primary concepts and details, classification of concepts, constructing hypotheses, generalizing, and drawing conclusions [30]. Science text is more difficult than other types of text because it contains domain-specific vocabulary and employs both causal and sequential text schemas; thus, high text comprehension skills are crucial to realize high science achievement [30,42].

To the best of our knowledge, few studies have correlated text comprehension and science achievement, although they do form a symbiotic relationship [29]. Previous studies have demonstrated that text comprehension skills have a strong effect on science achievement [29,30,42]. Text comprehension accounted for 17% of the variance in science achievement [28], and the correlation coefficient between science achievement and text comprehension was greater than 0.24 [43]. A study compiled several factors influencing science achievement using SEM, and they concluded that, after prior knowledge, text comprehension was the second largest contributing factor [42]. However, in a different study, a weak correlation was found between science achievement and specific text comprehension indicators (i.e., understanding vocabulary, noting detail, predicting outcomes, and making inferences) [30].

2.4. SES in mathematical problem-solving skills

SES plays a crucial role in the health, well-being, socioemotional development, and academic success of children [26]. The impact of COVID-19 on teaching and learning resulted in widespread remote learning, which raised an issue regarding inequalities in academic achievement due to different SES [27,44]. SES factors include the parents’ educational level, family income, parent involvement, and the parents’ occupation [1,25,26]. Family income and occupation are related to the social and economic resources available to the student at a given time, and the parents' education level is a relatively stable aspect [26]. In this study, we focused on only the parents' educational level and family income because these are the first indicators of SES used in the psychology field [27].

It has been found that the parents' educational level influences both mathematics achievement and mathematical problem-solving skills [25,26,45,46]. The level of education of parents is related to the parents' beliefs and behaviors, for example, engaging in a good interactions, providing quality academic support, and communicating high expectations [25,27,47]. In addition, parents can transfer their intelligence and knowledge up to 10% through genetic, and up to 30% through environmental pathways [27].

Parents' educational level has an important role in predicting students' mathematical problem-solving [26]. However, the studies that investigated how parents' educational level affects mathematics achievement and mathematical problem-solving skills still have shown varied findings [1,20,26,46]. For example, one study showed that parents' educational level has an indirect association with mathematics achievement, with parental expectation and total achievement as mediators [25,26]. Several other studies have confirmed that parents' educational level are strongly correlated with mathematics achievement with a correlation coefficient greater than 0.40 [45,46]. In addition, another study concluded that parents' educational level is weakly correlated with mathematics achievement with a correlation coefficient of less than 0.20 [20]. However, different countries can show different results, for example, in America, the parents' educational level is associated with mathematics achievement; however, this is not observed in Hong Kong [48]. Different roles can also affect different results, and studies have revealed that the father's educational level has an impact on mathematics achievement but the mother's educational level does not [1,47]. Since parents' educational level in mathematical problem-solving has a crucial role but the former studies still have mixed results, including parents' educational level as one of the factors that influences mathematical problem-solving skills is necessary to demonstrate the relevance of this factor in the current study.

A study conducted for grades 5–11 revealed that the mother's educational level is associated with problem-solving skills across grades, and the father's level of education is only significant in grades 7 and 8 [8]. However, another result concluded that there is no association between the parents' educational level and mathematical problem-solving skills [2].

Several previous studies have reported the parents' educational level as a single variable rather than dividing it into two related variables, that is, the educational levels of both the mother and the father [25,46,48]. This means that there is a symmetric correlation between the education levels of the mother and father.

Typically, the parents' educational level drives their occupation and income [25,27]. A study constructed a model in which the parents' educational level influences the parents' occupation and family income, and they proved that parents' educational level influenced family income [26,27].

Family income is also associated with mathematics achievement [23,26,46]. It has been found that family income affects students achievement because it is related to the provision of material resources and is associated with family stress, which influences parenting [27]. Family income also has a positive effect on mathematics skills ($r = 0.268$) [46] that is greater than the parents' level of education [23]. An indirect relationship has also been found between family income and mathematics skills with parent expectation as a mediator [26]. However, another study stated that there is no relation between family income and mathematics skills [25].

2.5. Theoretical model of factors influencing mathematical problem-solving

Several factors affect mathematical problem-solving skills in the science context, including cognitive and SES factors. The cognitive factors include mathematics DSPK, science background knowledge, and text comprehension skills, and the SES factors cover the parents' educational levels and family income. These variables directly affect mathematical problem-solving skills. Text comprehension skills also influence science knowledge. Thus, science background knowledge can function as a mediator in the relationship between text comprehension skills and mathematical problem-solving skills. In addition, the parents' educational levels are correlated symmetrically, and they drive family income. As a result, family income can be a mediator for the association between the parents' educational level and mathematical problem-solving skills. Fig. 1 illustrates a theoretical model of the factors affecting mathematical problem-solving skills.

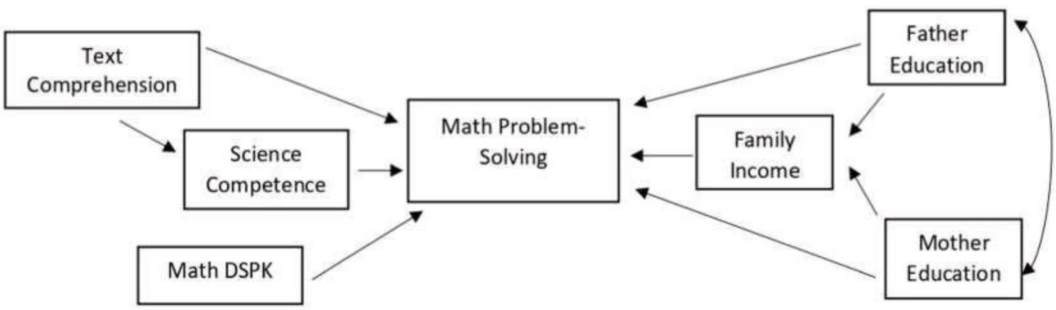


Fig. 1. Theoretical model of factors influencing mathematical problem-solving skills.

3. Method

3.1. Instrumentation

3.1.1. Science-related mathematical problem-solving test

The science-related mathematical problem-solving test is a scenario-based essay test with a mathematical scenario problem related to scientific phenomena. This test applies engineering based-design in the process of solving a problem using technology as a tool (e.g., calculator, grid paper, etc.). The test was developed under the integrated STEM-based mathematical problem-solving framework. It is a framework that applied the Program for International Student Assessment (PISA) creative problem-solving framework (exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting) as its core and combined other frameworks in every STEM discipline to construct indicators (i.e., scientific inquiry, engineering-based design, mathematics experimentation, and inquiry-based mathematics).

The test is targeted on grades seven to nine. In this test, there are three scenarios related to environmental management with one to two anchor scenarios for each grade. The total number of scenarios is six, i.e., “eco-friendly packaging” regarding designing an eco-friendly packaging and deciding the cheapest price (grade seven), “school park” regarding designing a planted school park with trees that can absorb the largest amount of CO2 (grade seven), “calorie vs. greenhouse gas emission” regarding designing a 1-day-menu that fulfills the nutritional requirements and has the lowest CO2 emission during their productions (grades seven to nine), “flood water reservoir” about designing a flood water reservoir and deciding the fastest time by pumps to absorb the average volume (grade eight), “city park” that has the similar idea with “school park” (grades eight and nine), and “infiltration well” regarding designing an infiltration well with complex constraints and determining the cost of its construction (grade nine). Each scenario involves a challenge pertaining to the topics of numbers and measurement, ratios and proportions, geometry, and statistics. Every scenario includes eight metacognitive prompt items to guide students to explore their problem-solving skills (e.g., the “Identifying the relevant and valuable given information” indicator includes the prompt “Mention the information that you need to answer the question!”).

The maximum score for each prompting item is five. A score of five indicates a complete and correct answer, four indicates a complete answer with a minor error, three indicates an incomplete but correct answer, two indicates an incomplete answer that contains an error, and one indicates a completely incorrect and irrelevant answer. The maximum score for each scenario is 40.

The psychometric evidence of the test was confirmed to be good and acceptable. The content validity index (CVI) value was greater than 0.67, the Cronbach Alpha was acceptable at .835, and the intraclass correlation coefficient (ICC) indicated moderate reliability (rxx = 0.628). All scenarios and prompting items were fit ($0.77 \leq$ weighted mean-square ≤ 1.59). In addition, discrimination values ($0.48 \leq$ discrimination value ≤ 0.93), the behavior of rating score, and reliabilities (scenarios’ reliability = 0.860; prompting items’ reliability = 0.938) were acceptable.

3.1.2. Mathematics DSPK test

This test which was developed based on the Indonesian mathematics curriculum 2013, includes 30 multiple choice questions. The test is a single test for all grades. It measures students’ conceptual and procedural mathematical knowledge. The construct of conceptual knowledge is the knowledge of meaning which assesses students’ understanding about basic mathematical concepts (e.g., the meaning of the fraction concept). The construct of procedural knowledge includes the integration of knowledge and the application of knowledge. Integration of knowledge targets interrelation among concepts in the understanding, planning, and evaluating phases (e. g., finding the correct statement regarding the problem-solving process). Application of knowledge focuses on applying knowledge in the application phase (e.g., calculating the reduction of CO2 emission).

The topics used are applied in the science-related mathematical problem-solving test and intersection topics in all graders based on the Indonesian curriculum 2013. The topics covered in this test include numbers (integers and fractions; nine questions) and measurement (four questions), ratios and proportions (eight questions), geometry (four questions), and statistics (five questions). Each topic involves four phases: understanding a basic concept (conceptual knowledge), interrelation among concepts in planning, executing, and evaluating (procedural knowledge). The test uses a science or social context. The test was designed as a set of inter-dependent multiple-choice items that represent a process, making it flexible to assess concepts and procedures (DSPK).

There are four options in every item with distractions. These distractions are designed based on the consequences of choosing the wrong answer in the previous item, as well as some common misconceptions, errors, and difficulties that students often encounter.

The test has a total score of 30, with 1 point awarded for the correct answer and 0 point for an incorrect answer. The psychometric evidence was confirmed to be good. The CVI values were greater than 0.83, and the ICC was acceptable (rxx = . 513). The construct of all items indicated fit ($0.90 \leq$ weighted mean-square ≤ 1.16) and were reliable ($\alpha = 0.74$) with various difficulty levels.

9. Text comprehension and science grades

The text comprehension grade was obtained from the diagnostic test scores for an Indonesian language literacy test. This language literacy test was a sample-based test for Indonesian students. At the beginning of an academic year, schools typically perform diagnostic assessments of students using a single test provided by Ministry of Education. The constructs of all the items indicated fit ($0.79 \leq$ weighted mean-square ≤ 1.43) and were reliable (EAP/PV reliability = .94). This test involves 21 items in finding information (7 items), interpreting and integrating the information (8 items), and evaluating and reflecting (6 items). There are three texts, namely narrative scientific text, scientific news text, and poetry text. In addition, this test involves simple and complex multiple-choice questions, short-answer questions, an essay, and matching type questions. The scores for this 60-min test are presented as percentages. Since the test is valid; hence students' grades from the test were used in the current study to assess the text comprehension skills.

The science grade was received from the mean score of the diagnostic tests scores relative to the environmental science topic provided by the books from the Indonesia's Ministry of Education. The tests are multiple-choice tests. Every school has to use the same tests within grade before learning every chapter. However, the tests are not the same test across grades, only have the similar competency indicators (e.g., the same test is used for grade 9 in all schools; however, the test for grade 9 is different from grade 7 despite similar indicators). The indicators are related to energy and photosynthesis (the plant organs involved in photosynthesis, the products of photosynthesis, factors affecting photosynthesis, experiments related to photosynthesis, the materials used in photosynthesis, and technologies inspired by the photosynthesis process), cause and effect of environmental issues and climate change (the greenhouse effect, methods to reduce greenhouse gas emissions, preventive actions against global warming, and the ozone layer), eco-friendly products (examples of eco-friendly products and their advantages), health and nutrition (nutrients, experiments related to nutrients, examples of foods containing vitamins and nutrients, and the functions of vitamins and nutrients in the body), biotechnology (examples of biotechnology products, the differences between conventional and modern biotechnology, and the effects of biotechnology) and scientific procedure (concepts of data, variables, and scientific procedures). There are 9 items in every test with four options. The mean score was converted into a percentage to facilitate uniformity.

The contents of the tests were confirmed valid and reliable with the CVI between 0.67 and 1. The results of the reliability of the tests for grades 7 to 9 were 0.98, 0.96, and 0.92, respectively.

3.1.3. SES questionnaire and document

The SES questionnaire covers the questions regarding parents' educational level both the father and mother (from elementary school to doctoral), family income (four categories based on the average level of the income in the province), and the number of siblings. The SES questionnaire was validated by teachers before administration. The information given is confirmed using the students' SES data from schools. The SES data from the schools are the official school data received from parents. Here, family income was grouped based on the total amount of the parents' incomes relative to the number of dependent children.

3.2. Participants

The population of the study included middle-school students (grades 7 to 9) from A-accreditation schools in both rural and urban areas in East Java, Indonesia. The Indonesia's Ministry of Education assesses the accreditation based on environment, human resources, and teaching and learning process. The highest accreditation is A-accreditation. There are 1886 A-accreditation schools in East Java, Indonesia, with a total of approximately 2 million students (npd.kemendikbud.go.id). Given the population, this study needs 1067 samples by applying a confidence level of 95% and a confidence interval of 3. Hence, we recruited 1067 participants and applied a stratified random sampling method. We chose schools in urban and rural areas and random classes in each grade. Table 1 describes the characteristics of the participants.

3.3. Data analysis

In this study, we investigated the effect of independent variables (i.e., mathematics DSPK, science background knowledge, text comprehension skills, parents' educational levels, and family income) on students' mathematical problem-solving skills. Here, we employed SEM with a maximum likelihood (ML) estimator, and the type of data was correlations and standard deviations. The SEM analysis was performed using the Mplus version 8.0 application.

To ensure the model fit, we investigated the Chi-squared test values, the comparative fit index (CFI), the root mean squared error of approximation (RMSEA), the Tucker–Lewis index (TLI), and the standardized root mean square residual (SRMR). The CFI and TLI

Table 1
Characteristics of participants.

Demographic characteristics		N	%
Gender	Boys	452	42.4
	Girls	615	57.6
Grade	7 (M age = 12.59, SD = 0.61)	380	35.61
	8 (M age = 13.42, SD = 0.59)	331	31.02
	9 (M age = 14.50, SD = 0.59)	356	33.36
School location	Rural	427	40.02
	Urban	640	59.98
Ethnic	Javanese	1015	95.01
	Madurese	35	3.3
	Batak	2	0.20
	Java-Madurese	3	0.29
	Sundanese	5	0.50
	Betawi	2	0.20
	Others	5	0.50

values were categorized as good if they were greater than or equal to 0.95; however, they are still acceptable with a value of approximately .90 [49]. In addition, a good SRMR value is less than 0.08 [49], and the RMSEA value was considered acceptable if it was less than 0.08 and marginal if it was in the range [0.08, 0.10] [50]. The model fit will be improved by creating a modified model. After we obtained a fit model, we examined the independent variables that influence mathematical problem-solving skills according to the significance values and correlation coefficients. In addition, we identified the standardized estimates in terms of their direct effect, indirect effect, and total effect for all independent variables.

3.4. Procedure

The data collection process was performed from July to September 2022. Note that the researchers applied for ethical approval from the institutional review board (IRB) of the university and were granted with the ethical approval number of 7/2022 prior to collecting the data. The ethical approval letter was given to several schools and they asked to response their participation. After confirming participation, the researchers discussed the data collection procedure with mathematics teachers.

On the first day, the students were asked to take the science-related mathematical problem-solving test for 3 h (with a break in every hour) using a paper-based technique. Then, on another day, the students used an online platform to take the mathematics DSPK test (90 min) and SES questionnaire (15 min). In addition, the students' grades for science and text comprehension were requested from the schools. These students' grades documents will be employed to assess students' science knowledge and text comprehension skills.

The students' answers for the science-related mathematical problem-solving test were scored by two raters, an author and a teacher who graduated with master's degrees in mathematics education, to ensure the consistency of the implementation of the rating scale. The obtained ICC value was 0.992, which indicates good and consistent ($\alpha = 0.996$). The data from tests, documents, and the questionnaire were used to calculate the correlations and standard deviations. The correlations and standard deviations data were then used to analyze the theoretical model of factors influencing mathematical problem-solving skills using SEM.

4. Results

4.1. Descriptive statistics

Before reporting the results of the descriptive statistics, we first describe the students' performance in the mathematical problem-solving task to demonstrate the variation in their performance levels. Students' performance in science-related mathematical problem-solving is categorized as average (M = 60.49, SD = 19.266). Overall, 52.1% of students scored below average. The students' performance based on the grade showed that grade 8 students achieved the highest mean score (M = 61.98, SD = 20.35), followed by students in grade 9 (M = 60.81, SD = 13.82) and grade 7 (M = 58.89, SD = 22.36). According to the level of the mothers' education, students with mothers holding a master's degree exhibited the highest performance in science-related mathematical problem-solving (M = 67.82, SD = 20.53), followed by students with mothers holding a doctoral degree (M = 63.28, SD = 11.84), bachelor's degree (M = 61.17, SD = 19.30), high-school degree (M = 61.10, SD = 19.09), middle-school degree (M = 58.84, SD = 20.06), and elementary school degree (M = 55.37, SD = 18.62). However, the pattern differed based on the fathers' educational level. Students with fathers holding a bachelor's degree performed better (M = 63.92, SD = 19.41) compared to those with fathers holding a master's degree (M = 62.79, SD = 19.55), doctoral degree (M = 61.46, SD = 22.52), high-school degree (M = 60.55, SD = 18.79), elementary school degree (M = 59.97, SD = 20.58), and middle school degree (M = 54.78, SD = 19.40). In addition, students from low family income demonstrated the lowest performance in science-related mathematical problem-solving (M = 59.36, SD = 19.00) compared with students from middle family income (M = 63.59, SD = 19.82) and high family income (M = 68.22, SD = 18.65).

In the following, we report the variation and association of the observed variables obtained from the given data through standard deviations and correlation coefficients to infer causality. Table 2 shows the correlation coefficients and standard deviations for each variable.

The standard deviations of the variables are large indicating that the data is quite spread out or there might be outliers. However, it is not necessary to remove the outlier data since it is not noise data. In addition, deleting the outlier does not influence the model due to the large sample size. All variables were correlated significantly, except the association between the father's educational level, the

Table 2
Correlation coefficients and standard deviations of variables.

Variables	1	2	3	4	5	6	7
1. Mathematical problem-solving skill	1						
2. Math DSPK	.668**	1					
3. Science knowledge	.590**	.362**	1				
4. Text comprehension skill	.498**	.376**	.294**	1			
5. Mother's education	.095**	.032	.108**	.056	1		
6. Father's education	.090**	.034	.092**	.033	.563**	1	
7. Family income	.113**	.017	.081**	.078*	.381**	.364**	1
SD	19.266	4.025	29.819	16.727	.944	.887	.505

**p < .01; *p < .05.

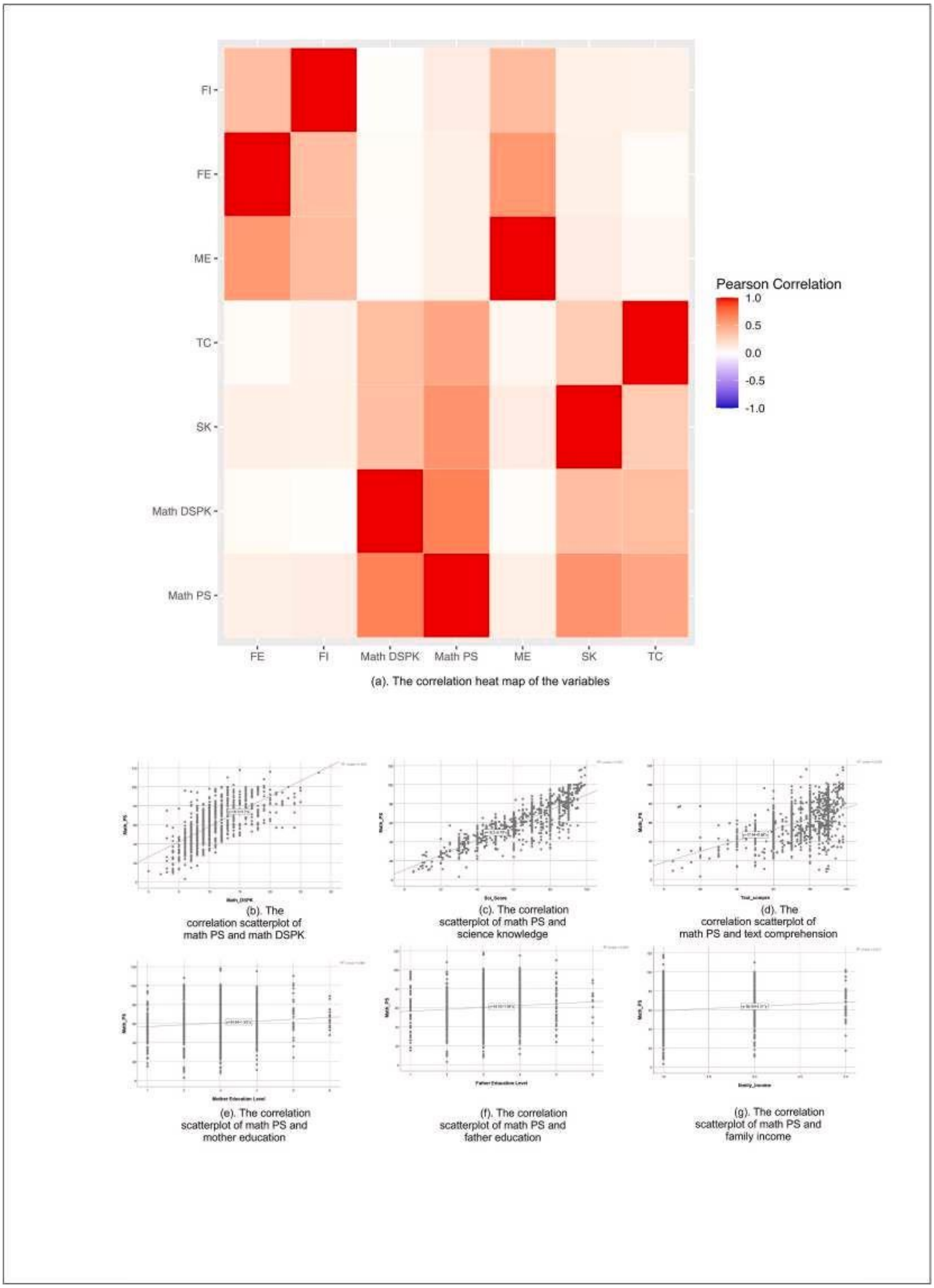


Fig. 2. (a) The correlation heat map of the variables; (b) The correlation scatterplot of mathematical problem-solving and mathematics DSPK; (c) The correlation scatterplot of mathematical problem-solving and science knowledge; (d) The correlation scatterplot of mathematical problem-solving and text comprehension; (e) The correlation scatterplot of mathematical problem-solving and mother education; (f) The correlation scatterplot of mathematical problem-solving and father education; (g) The correlation scatterplot of mathematical problem-solving and family income.

mother's educational level, and family income and mathematics DSPK. In addition, no significant correlation was observed between the father or mother's educational level and text comprehension skills. The strongest correlation was found between mathematical problem-solving skills and mathematics DSPK. We also found weak correlation coefficients or correlation coefficient less than 0.4, which may cause a problem of insignificant association in the loading of the measurement model. For example, the correlation coefficient between father's educational level and mathematical problem-solving skills was 0.090. Fig. 2 represents a correlation heat map and correlation scatterplots of the variables.

4.2. Goodness of fit indices

In the following, we discuss five goodness of fit indices, i.e., χ^2, CFI, TLI, RMSEA, and SRMR. We found that the CFI, TLI, and SRMR results are acceptable. Although the TLI is less than 0.90, but it is relatively close to 0.90. The TLI value close to 0.90 indicated that the model is fit, but needed improvement since it was less than 0.90. The RMSEA result is considered marginal. According to the results for several of the goodness of fit indices, the model was categorized as a fit model; however, it was close to the minimum criteria of TLI, CFI, RMSEA, and SRMR.

Thus, we attempted to improve the results in terms of the goodness of fit indices by producing Model 2, which does not include a direct path from the parents' education level to mathematical problem-solving. In the first model, we obtained a nonsignificant p value in the direct path of the parents' education level to mathematical problem-solving. Here, the CFI, TLI, RMSEA, and SRMR results are acceptable. Even though the RMSEA value for Model 2 was close to 0.08, we found that Model 2 obtained better goodness of fit results. Table 3 represents the results of the goodness of fit indices of Model 1 and Model 2.

4.3. Estimates of structural model

Theoretical Model 1 was analyzed, and we found that two independent variables were not significantly associated with mathematical problem-solving skills, i.e., the father's and mother's educational levels. This means that the parents' educational levels do not directly influence mathematical problem-solving skills. The most influential variable was mathematics DSPK ($r = 0.478$). We also found that all cognitive factors strongly influenced mathematical problem-solving skills rather than the SES factors. With Model 2, we found that all independent variables significantly affected mathematical problem-solving skills, where mathematics DSPK ($r = 0.478$) remained the most influential variable. Table 4 shows the standardized estimates of the structural model of Models 1 and 2.

4.4. Mediating effect of structural model

The results for both Models 1 and 2 indicated that mathematics DSPK had the largest total effect, followed by science background knowledge and text comprehension skills. According to theoretical Models 1 and 2, text comprehension skills influence mathematical problem-solving mediated by science knowledge. This means that text comprehension skills can indirectly influence mathematical problem-solving through science knowledge. The results for Model 1 demonstrated that text comprehension skills affected mathematical problem-solving skills both directly ($r = 0.228$) and indirectly ($r = 0.109$). However, the direct effect was greater than the indirect effect. A similar result was obtained with Model 2, where the direct effect of text comprehension skills was stronger than the indirect effect. Although the total effect of text comprehension skills on mathematical problem-solving skills in Models 1 and 2 was the same, in Model 1, the direct effect was greater, and the indirect effect was less than that of Model 2.

The results obtained for Model 1 indicated that the parents' educational levels do not significantly or directly affect mathematical problem-solving skills. However, both of these variables influenced mathematical problem-solving skills indirectly through family income, but the effects were weak ($r_{motheredu} = 0.014$ and $r_{fatheredu} = 0.012$). For Model 2, when the direct association between mathematical problem-solving skills and the parents' educational levels were removed, the indirect effects of both variables on mathematical problem-solving skills were stronger. The effect of family income was also stronger in Model 2. Table 5 shows the mediating effect of the structural Models 1 and 2 (Figs. 3 and 4 show the standardized coefficients of Models 1 and 2).

5. Discussion

Mathematical problem-solving in the science context has been identified as the most difficult skill for school age students because such problems are typically presented using natural language; thus, students must be able to understand the meaning of the problem's text prior to applying mathematical concepts [12–15]. In addition, science is integrated as a context of the mathematical problem;

Table 3
The results of the goodness of fit indices of model 1 and model 2.

	Model 1	Model 2
χ^2	$\chi^2(10) = 110.435, p < .001$	$\chi^2(12) = 110.994, p < .001$
CFI	.931	.932
TLI	.897	.915
RMSEA	.097	.088
SRMR	.076	.076

Table 4
Standardized estimate of structural model of Models 1 and 2.

Endogenous variable	Exploratory variable	Model 1			Model 2		
		Standardized estimate	S.E.	P	Standardized estimate	S.E.	p
Mathematical problem-solving skill	Mathematics DSPK	.478	.020	<.001	.478	.020	<.001
	Science knowledge	.372	.020	<.001	.373	.020	<.001
	Text comprehension	.228	.022	<.001	.227	.022	<.001
	Mother's education	.001	.024	.955	–	–	–
	Father's education	.015	.024	.533	–	–	–
	Family income	.056	.022	.009	.062	.020	.001

Note: "-" indicates that there is no path between these variables.

Table 5
Mediating effect of structural models 1 and 2.

Endogenous variable	Exploratory variable	Model 1			Model 2		
		Direct effect	Indirect effect	Total effect	Direct effect	Indirect effect	Total effect
Mathematical problem-solving skill	Mathematics DSPK	.478*	–	.478*	.478*	–	.478*
	Science competencies	.372*	–	.372*	.373*	–	.373*
	Text comprehension	.228*	.109*	.337*	.227*	.110*	.337*
	Mother's education	.001	.014*	.015*	–	.016*	.014*
	Father's education	.015	.012*	.027*	–	.014*	.016*
	Family income	.056*	–	.056*	.062*	–	.062*

Note: "-" indicates that there is no path between these variables.
*p < .01.

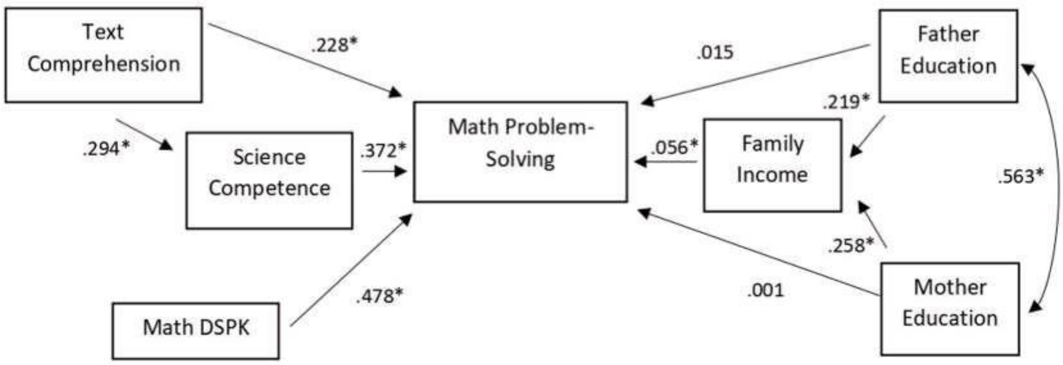


Fig. 3. Standardized coefficients of Model 1.

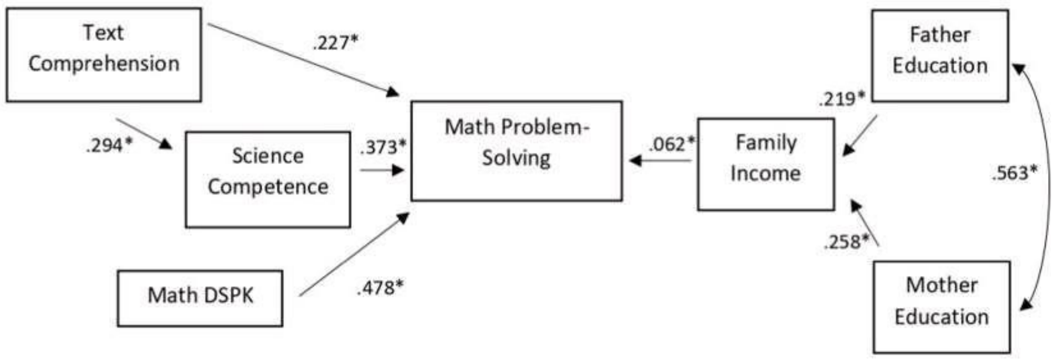


Fig. 4. Standardized coefficients of Model 2.

thus, students should have sufficient science knowledge to be successful in solving it [6,9]. Therefore, success in mathematical problem-solving is obtained through a strong mathematics DSPK and knowledge in both science and text comprehension skills [6,9,12, 14].

In addition to cognitive factors, the SES factor also contributes to success in mathematical problem-solving [25–27]. The first indicators of SES in psychology research were the parents' educational level and family income [27]. These factors have been shown to contribute to success in mathematical problem-solving [23,26,46,47]. The parents' educational level leads to positive behaviors and beliefs [25,27,47], and family income affects mathematical problem-solving skill through the provision of material resources [25,27].

In this study, we constructed a theoretical model of factors that affect mathematical problem-solving skills, called Model 1. In this model, the exploratory variables included mathematics DSPK, science background knowledge, text comprehension skills, the mother's educational level, the father's educational level, and family income. In this theoretical model, these variables had a direct association.

However, we also found that text comprehension also had an indirect effect through science knowledge. Understanding a science problem requires text comprehension skills because science problems contain words and involve unique vocabulary [28–30,43]. We also found that the parents' educational levels have an indirect effect through family income in the theoretical model because higher educational levels tend to lead to higher incomes [25,27].

The goodness fit results for Model 1 were acceptable; however, the RMSEA value was marginal, and the TLI value was close to the cut-off point which is 0.90 [49,50]. Overall, Model 1 was acceptable; however, improvements were required. A model can be modified by adding parameters and paths to increase model complexity or remove unnecessary parameters and paths to create a simpler model [51]. In this study, we decided to remove the direct association between the father and mother's educational levels and mathematical problem-solving skills because there were no significant effects in the model. The modified model, i.e., Model 2, obtained better CFI, TLI, RMSEA, and SRMR results, which indicated that Model 2 exhibited acceptable goodness of fit [49,50].

The largest correlation coefficient was between mathematical problem-solving skills and mathematics DSPK ($r = 0.684$). It means that mathematics DSPK is the strongest factor influencing mathematical problem-solving skills ($r = 0.478$). This result was similar to results reported by previous studies, i.e., mathematics knowledge is strongly correlated to problem-solving, and the correlation coefficient was greater than 0.40 [6,35,36,38]. In mathematical problem-solving, students must utilize mathematics knowledge in the knowledge application phase; thus, prior mathematics knowledge is a crucial factor to realize successful mathematical problem-solving skills [8,19].

Science background knowledge was the second strongest factor affecting problem-solving skills. ($r = 0.373$), and this result is in agreement with a study's findings [6]. In addition, in mathematical problem-solving in the science context, science knowledge contributes in the application of knowledge and the performance of the inquiry process [9]. Thus, the notion that science competencies influence mathematical problem-solving in the science context was predicted effectively. However, the correlation coefficient of science knowledge and mathematical problem-solving in the current study is weaker than that observed in previous studies, although the correlation is still relatively high [6,9].

We found that text comprehension skills have both direct and indirect effects on mathematical problem-solving skills, and these results confirmed the finding of previous studies [3,14,40,41]. Mathematical problem-solving is typically presented using natural language, and to solve problems, students must understand the information and build mental representations using text comprehension skills [14,32]. However, the correlation coefficient of the text comprehension skills and mathematical problem-solving skills in the present study is weaker than that reported in previous relevant studies. According to the results, text comprehension also affected science competencies, which is also in line with the results of previous studies because a science problem presented in linguistic form requires students to comprehend the text to solve it [29,30,42,43]. An indirect effect of text comprehension skills on mathematical problem-solving skills was identified with science competencies functioning as a mediator ($r = 0.110$). This indirect effect has not been considered in previous studies.

The results revealed that the correlation coefficients of the parents' educational levels on mathematical problem-solving were not significant. In Model 1, the parents' educational levels did not significantly or directly affect mathematical problem-solving skills. It is in accordance with the results of previous studies [2,48] and contradicts with the majority results [25,26,45,46]. The culture and social interaction of different countries affect the results of the relationship between parents' educational level and mathematical problem-solving skills [48]. We hypothesize that the nonsignificant correlation of parents' educational level and mathematical problem-solving is due to (1) the trend of full working time for parents in Indonesia; and (2) well-educated parents are not always having economically well-being. Parents, both the mother and father, with high levels of educational tend to work from 9 a.m. to 5 p.m. (or more); thus, they have less time to focus on their children. The second hypothesis is related to the change in work skills to contemporary skills. Thus, a high level of educational will not guarantee a better job. However, we must confirm these hypotheses. In Model 1, there is no direct effect from the parents' educational level; thus, we constructed Model 2, which did not include a direct relationship between the parents' educational levels and mathematical problem-solving skills.

This phenomenon contributed to the results reported in previous studies, i.e., there is another factor that mediates the association between the parents' educational levels and mathematical problem-solving skills [2,25,26]. According to the empirical results of Models 1 and 2, the results agreed with the previous studies that found the mother and father's educational levels influenced mathematical problem-solving skills indirectly [25,26]. Here, the mediating variable was family income, which is different from previous studies. We found that family income directly affects mathematical problem-solving skill, which is in agreement with the results of several studies [23,46], but contradicts one [25]. However, the effect was weak compared to the former studies. In Model 1, the direct effect of family income was less than that in Model 2 because in Model 2 we removed the paths of nonsignificant variables.

6. Conclusion

Mathematical problem-solving in the science context involves complex skills, both mathematical and nonmathematical skills. In this study, we constructed two theoretical models, i.e., the original model and a modified model (Models 1 and 2, respectively), to analyze the factors that influence mathematical problem-solving skills. We found that both models obtained acceptable results; however, according to the results, the modified Model 2 exhibited better goodness of fit indices.

This study confirmed several cognitive and SES factors that influence mathematical problem-solving skills, and we identified both direct and indirect effects. We found cognitive factors have a greater effect than SES factors. The cognitive factors that directly influence mathematical problem-solving were mathematics DSPK, science background knowledge, and text comprehension skills, with mathematical DSPK being the factor with the greatest influence. Text comprehension skills were also found to indirectly affect mathematical problem-solving skills through science knowledge.

Family income was only the SES factor that directly affected mathematical problem-solving skills. The parents' educational levels were found to have an indirect effect on mathematical problem-solving skills through family income. The reasons regarding less time spending with children and economic well-being were still needed to be confirmed.

This study has limitations relative to the limited number of variables in the SES factors, the data source or instruments, and the large standard deviation. In this study, we only covered two SES indicators; however, there are several other SES indicators are known to affect mathematical problem-solving skills (e.g., parents' occupation and family structure). These potential unmeasured variables are predicted to have impact on students' mathematical problem-solving skills. The second limitation was that the score of science knowledge was from the score of diagnostic tests that were conducted not using a single test, but only using the tests that have the same constructs, contents, and indicators of competencies. In addition, the categories of parents' educational levels are relatively general, and they may not adequately reflect the complex facets of parents' educational experiences, such as the disciplines they studied or the depth of their mathematical knowledge. These aspects may influence parental support and students' exposure to mathematical concepts at home. Hence, it is considered using more precise instrument for assessing parents' education. The third limitation was that the variables have large standard deviations, which require further investigation.

Despite these limitations, the results of this study can be used by teachers to consider the factors that strongly affect mathematical problem-solving. Thus, teachers can better prepare appropriate teaching and learning strategies that consider and include these factors. Teachers should consider mathematical DSPK and science knowledge before introducing complex subject matter, e.g., mathematics problem-solving in the science context. Since mathematics DSPK is the strongest factor influencing mathematical problem-solving, it is suggested to (1) recall students' mathematics prior knowledge before mathematical problem-solving activity; (2) enhance students' mathematics prior knowledge in the most difficult knowledge and mathematics topics; and (3) design a comprehensive teaching and learning approach to improve students' prior knowledge when such knowledge is needed but appears to be poor. In addition, to consider science knowledge and text comprehension skills factors, it is necessary to familiarize students with nonroutine word problems in the science context using a constructivist teaching method (e.g., mathematics inquiry-based learning).

Future study is wide open for researchers in relevant fields. For example, the present study was conducted in a large sample size; hence the results can serve as a basis for relevant studies aimed at determining factors that influence mathematical problem-solving skills. However, the generalization of the results of this study is limited to the Indonesian context. The theoretical model examined in this study can be adapted or replicated for relevant studies to be conducted in different demographic backgrounds (e.g., different countries). This adaptability is possible because the theoretical model is formulated based on previous studies from several different contexts. The analysis of the adapted model in the sample from different countries will create a strong basis for generalization regarding the factors impacting mathematical problem-solving skills. In addition, it is possible to consider additional variables in terms of the SES factors.

Author contribution statement

Ijtihadi Kamilia Amalina: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Tibor Vidákovich: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by the University of Szeged Open Access Fund with the grant number of 6007.

Data availability statement

Data will be made available on request.

Additional information

No additional information is available for this paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

[1] A. Imam, G.P. Singh, Influence of gender, parental education, and parental occupation on mathematics achievement of secondary school students, Indian J. Res. 4 (11) (2015) 187–190 [Online]. Available: https://www.worldwidejournals.com/paripex/recent_issues_pdf/2015/November/November_2015_1448104482_63.pdf.

[2] B. Guven, B.O. Cabakcor, Factors influencing mathematical problem-solving achievement of seventh grade Turkish students, Learn. Indiv Differ 23 (1) (2013) 131–137, https://doi.org/10.1016/j.lindif.2012.10.003.

[3] S. Anjum, Gender difference in mathematics achievement and its relation with reading comprehension of children at upper primary stage, J. Educ. Pract. 6 (16) (2015) 71–75 [Online]. Available: https://files.eric.ed.gov/fulltext/EJ1079951.pdf.

[4] Y. Nakakoji, R. Wilson, Interdisciplinary learning in mathematics and science: transfer of learning for 21st century problem solving at university, J. Intell. 8 (3) (2020) 1–23, https://doi.org/10.3390/jintelligence8030032.

[5] S. Turs¸ucu, J. Spandaw, M.J. de Vries, The effectiveness of activation of prior mathematical knowledge during problem-solving in physics, Eurasia J. Math. Sci. Technol. Educ. 16 (4) (2020), https://doi.org/10.29333/ejmste/116446.

[6] B. Csapo´, G. Moln´ar, Potential for assessing dynamic problem-solving at the beginning of higher education studies, Front. Psychol. 8 (2017), https://doi.org/10.3389/fpsyg.2017.02022.

[7] A. Vista, The role of reading comprehension in maths achievement growth: investigating the magnitude and mechanism of the mediating effect on maths achievement in australian classrooms, Int. J. Educ. Res. 62 (2013) 21–35, https://doi.org/10.1016/j.ijer.2013.06.009.

[8] S. Greiff, S. Wüstenberg, G. Moln´ar, A. Fischer, J. Funke, B. Csapo´, Complex problem solving in educational contexts-something beyond g: concept, assessment, measurement invariance, and construct validity, J. Educ. Psychol. 105 (2) (2013) 364–379, https://doi.org/10.1037/a0031856.

[9] R. Scherer, J.F. Beckmann, The acquisition of problem solving competence: evidence from 41 countries that math and science education matters, Large-Scale Assess. Educ. 2 (1) (2014) 1–22, https://doi.org/10.1186/s40536-014-0010-7.

[10] F.K. Lester, Thoughts about research on mathematical problem- solving, Math. Enthous. 10 (1) (2013) 245–278, https://doi.org/10.54870/1551-3440.1267.

[11] K. Nunokawa, Mathematical problem solving and learning mathematics: what we expect students to obtain, J. Math. Behav. 24 (3–4) (2005) 325–340, https://doi.org/10.1016/j.jmathb.2005.09.002.

[12] G. Daroczy, D. Meurers, J. Heller, M. Wolska, H.C. Nürk, The interaction of linguistic and arithmetic factors affects adult performance on arithmetic word problems, Cognit. Process. 21 (1) (2020) 105–125, https://doi.org/10.1007/s10339-019-00948-5.

[13] O.A. Imam, M.A. Mastura, H. Jamil, Correlation between reading comprehension skills and students' performance in mathematics, Int. J. Eval. Res. Educ. 2 (1) (2013), https://doi.org/10.11591/ijere.v2i1.1803.

[14] M. Spencer, L.S. Fuchs, D. Fuchs, Language-related longitudinal predictors of arithmetic word problem solving: a structural equation modeling approach, Contemp. Educ. Psychol. 60 (November 2019) (2020), 101825, https://doi.org/10.1016/j.cedpsych.2019.101825.

[15] A.J.H. Boonen, F. Van Wesel, J. Jolles, M. Van der Schoot, The role of visual representation type, spatial ability, and reading comprehension in word problem solving: an item-level analysis in elementary school children, Int. J. Educ. Res. 68 (2014) 15–26, https://doi.org/10.1016/j.ijer.2014.08.001.

[16] M.P. Carlson, I. Bloom, The cyclic nature of problem solving: an emergent multidimensional problem-solving framework, Educ. Stud. Math. 58 (1) (2005) 45–75, https://doi.org/10.1007/s10649-005-0808-x.

[17] A.H. Schoenfeld, Learning to think mathematically: problem solving, metacognition, and sense making in mathematics (reprint), J. Educ. 196 (2) (2016) 1–38, https://doi.org/10.1177/002205741619600202.

[18] L.M. Soto-Ardila, A. Caballero-Carrasco, L.M. Casas-García, Teacher expectations and students' achievement in solving elementary arithmetic problems, Heliyon 8 (5) (2022), e09447, https://doi.org/10.1016/j.heliyon.2022.e09447.

[19] N.J.D. Mills, ALEKS constructs as predictors of high school mathematics achievement for struggling students, Heliyon 7 (6) (2021), e07345, https://doi.org/10.1016/j.heliyon.2021.e07345.

[20] K.P. Pangeni, Factors determining educational quality: student mathematics achievement in Nepal, Int. J. Educ. Dev. 34 (1) (2014) 30–41, https://doi.org/10.1016/j.ijedudev.2013.03.001.

[21] A. Bahar, The Influence of Cognitive Abilities on Mathematical Problem Solving Performance, The University of Arizona, Arizona, 2013.

[22] M. O¨sterholm, E. Bergqvist, What is so special about mathematical texts? Analyses of common claims in research literature and of properties of textbooks, ZDM - Int. J. Math. Educ. 45 (5) (2013) 751–763, https://doi.org/10.1007/s11858-013-0522-6.

[23] G.N. Marks, A. Pokropek, Family income effects on mathematics achievement: their relative magnitude and causal pathways, Oxf. Rev. Educ. 45 (6) (2019) 769–785, https://doi.org/10.1080/03054985.2019.1620717.

[24] L. Wang, X. Li, N. Li, Socio-economic status and mathematics achievement in China: a review, ZDM - Math. Educ. 46 (7) (2014) 1051–1060, https://doi.org/10.1007/s11858-014-0617-8.

[25] B.O. Alomar, Personal and family factors as predictors of pupils' mathematics achievement, Psychol. Rep. 101 (1) (2007) 259–269, https://doi.org/10.2466/PR0.101.1.259-269.

[26] M. Hasco¨et, V. Giaconi, L. Jamain, Family socioeconomic status and parental expectations affect mathematics achievement in a national sample of Chilean students, Int. J. Behav. Dev. 45 (2) (2021) 122–132, https://doi.org/10.1177/0165025420965731.

[27] P.E. Davis-Kean, L.A. Tighe, N.E. Waters, The role of parent educational attainment in parenting and children's development, Curr. Dir. Psychol. Sci. 30 (2) (2021) 186–192, https://doi.org/10.1177/0963721421993116.

[28] N. Bayat, G. S¸ekerciog˘lu, S. Bakir, The relationship between reading comprehension and success in science, Egit. ve Bilim 39 (176) (2014) 457–466, https://doi.org/10.15390/EB.2014.3693.

[29] F. Cano, A´. García, A.B.G. Berb´en, F. Justicia, Science learning: a path analysis of its links with reading comprehension, question-asking in class and science achievement, Int. J. Sci. Educ. 36 (10) (2014) 1710–1732, https://doi.org/10.1080/09500693.2013.876678.

[30] O.A. Imam, et al., Reading comprehension skills and performance in science among high school, Asia Pac. J. Educ. Educ. 29 (2014) 81–94 [Online]. Available: http://eprints.usm.my/34754/1/Art_5(81-94).pdf.

[31] G. Moln´ar, S. Greiff, B. Csapo´, Inductive reasoning, domain specific and complex problem solving: relations and development, Think. Skills Creativ. 9 (2013) 35–45, https://doi.org/10.1016/j.tsc.2013.03.002.

[32] X. Lin, Investigating the unique predictors of word-problem solving using meta-analytic structural equation modeling, Educ. Psychol. Rev. 33 (3) (2021) 1097–1124, https://doi.org/10.1007/s10648-020-09554-w.

[33] I. V Silao, Factors affecting the mathematics problem solving skills of Filipino pupils, Int. J. Sci. Res. Publ. 8 (2) (2018) 487 [Online]. Available: www.ijsrp.org.

[34] H.M. Süß, A. Kretzschmar, Impact of cognitive abilities and prior knowledge on complex problem solving performance - empirical results and a plea for ecologically valid microworlds, Front. Psychol. 9 (MAY) (2018) 1–22, https://doi.org/10.3389/fpsyg.2018.00626.

[35] J.L. Booth, J.L. Davenport, The role of problem representation and feature knowledge in algebraic equation-solving, J. Math. Behav. 32 (3) (2013) 415–423, https://doi.org/10.1016/j.jmathb.2013.04.003.

[36] A.K. Jitendra, A.E. Lein, J.R. Star, D.N. Dupuis, The contribution of domain-specific knowledge in predicting students' proportional word problem-solving performance, Educ. Res. Eval. 19 (8) (Nov. 2013) 700–716, https://doi.org/10.1080/13803611.2013.845107.

[37] C. Oksuz, Association of domain-specific knowledge and analytical ability with insight problem solving in mathematics, Int. J. Pedagog. Learn. 5 (1) (2009) 138–153, https://doi.org/10.5172/ijpl.5.1.138.
[38] M.A. Al-Mutawah, R. Thomas, A. Eid, E.Y. Mahmoud, M.J. Fateel, Conceptual understanding, procedural knowledge and problem-solving skills in mathematics: high school graduates work analysis and standpoints, Int. J. Educ. Pract. 7 (3) (2019) 258–273, https://doi.org/10.18488/journal.61.2019.73.258.273.
[39] P.M. Vilenius-Tuohimaa, K. Aunola, J.E. Nurmi, The association between mathematical word problems and reading comprehension, Educ. Psychol. 28 (4) (2008) 409–426, https://doi.org/10.1080/01443410701708228.
[40] S. Akbasli, M. Sahin, Z. Yaykiran, The effect of reading comprehension on the performance in science and mathematics, J. Educ. Pract. 7 (16) (2016) 108–121 [Online]. Available: https://files.eric.ed.gov/fulltext/EJ1108657.pdf.
[41] L. Salihu, M. Aro, P. R¨as¨anen, Children with learning difficulties in mathematics: relating mathematics skills and reading comprehension, Issues Educ. Res. 28 (4) (2018) 1024–1038 [Online]. Available: https://www.semanticscholar.org/paper/Children-with-learning-difficulties-in-mathematics%3A-Salihu-Aro/5a9c7f13502ebc431d07f769331763030cfdcd4e.
[42] J.G. Cromley, L.E. Snyder-Hogan, U.A. Luciw-Dubas, Reading comprehension of scientific text: a domain-specific test of the direct and inferential mediation model of reading comprehension, J. Educ. Psychol. 102 (3) (2010) 687–700, https://doi.org/10.1037/a0019452.
[43] T. O'Reilly, D.S. McNamara, The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional 'high-stakes' measures of high school students' science achievement, Am. Educ. Res. J. 44 (1) (2007) 161–196, https://doi.org/10.3102/0002831206298171.
[44] A. Mukuka, O. Shumba, H.M. Mulenga, Students' experiences with remote learning during the COVID-19 school closure: implications for mathematics education, Heliyon 7 (7) (2021), e07523, https://doi.org/10.1016/j.heliyon.2021.e07523.
[45] A. Kodippili, Parents' education level in students' mathematics achievement; do school factors matter? Acad. Leader. 9 (1) (2011) [Online]. Available: https://scholars.fhsu.edu/alj/vol9/iss1/39/.
[46] Y. Alghazo, R. Alghazo, The effect of parental involvement and socioeconomic status on elementary students' mathematics achievement, J. Soc. Sci. Humanit. 1 (5) (2015) 521–527 [Online]. Available: http://www.aiscience.org/journal/paperInfo/jssh?paperId=2312.
[47] C. Engin-Demir, Factors influencing the academic achievement of the Turkish urban poor, Int. J. Educ. Dev. 29 (1) (2009) 17–29, https://doi.org/10.1016/j.ijedudev.2008.03.003.
[48] D.B. Wang, Family background factors and mathematics success: a comparison of Chinese and US students, Int. J. Educ. Res. 41 (1 SPEC) (2004) 40–54, https://doi.org/10.1016/j.ijer.2005.04.013.
[49] L.T. Hu, P.M. Bentler, Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives, Struct. Equ. Model. 6 (1) (1999) 1–55, https://doi.org/10.1080/10705519909540118.
[50] L.R. Fabrigar, D.T. Wegener, R.C. MacCallum, E.J. Strahan, Evaluating the use of exploratory factor analysis in psychological research, Psychol. Methods 4 (3) (1999) 272–299. https://psycnet.apa.org/doi/10.1037/1082-989X.4.3.272.
[51] H. Kang, J.W. Ahn, Model setting and interpretation of results in research using structural equation modeling: a checklist with guiding questions for reporting, Asian Nurs. Res. (Korean. Soc. Nurs. Sci). 15 (3) (2021) 157–162, https://doi.org/10.1016/j.anr.2021.06.001.

CHAPTER 3. DISCUSSION, CONCLUSIONS, AND EDUCATIONAL IMPLICATIONS

3.1 Discussion and Conclusions

3.1.1 STEM Problem-Solving Assessment and Tool Development

The study assessed Indonesian middle school students' mathematical problem-solving skills and examined the cognitive and SES factors influencing them. The task was designed using an interdisciplinary STEM-based context. Given the complexity of assessment in STEM, it is required to review which assessment tools and frameworks are appropriate for different STEM disciplines. However, there is a lack of comprehensive reviews, specifically addressing assessments and frameworks in STEM problem-solving (Gao et al., 2020).

Thus, the assessment of STEM problem-solving was comprehensively and systematically reviewed using the PRISMA method in this study. It focused on trends in participants, disciplines, assessment tools (including their drawbacks and advantages), and the framework used. It revealed that STEM problem-solving among middle school students remains a relevant research topic, as most studies were published between 2016 and 2020 focusing on middle school participants. However, interdisciplinary problem-solving, particularly in Indonesia, remains underexplored. These findings agree with previous studies that STEM assessment focused on monodisciplinary assessment within middle-school students, as they begin to think critically (Gao et al., 2020; Kelley & Knowles, 2016). However, previous studies reviewed the assessment in general STEM without specifying on problem-solving skills, although similar findings were highlighted.

Different trends were observed in the assessment tools used in each discipline. Closed-ended essay questions were common for assessing cognitive monodisciplinary problem-solving, while complex scenario-based essays were prevalent for interdisciplinary problem-solving assessment. Common challenges included the unavailability of integrated frameworks, time-consuming assessments, and insufficient psychometric evidence. They used fundamental frameworks such as heuristic and PISA creative problem-solving frameworks for mathematics, scientific inquiry for science, and engineering-based design for engineering. However, a standardized and clearly integrated framework was unavailable to measure interdisciplinary STEM problem-solving.

To address the needs identified in the review, an integrated STEM-based mathematical problem-solving test was developed and validated for Indonesian middle school students. This test solves previous challenges related to the complexity and classroom applicability of existing tests, their emphasis on product-oriented outcomes, and the underrepresentation of mathematics (Annaggar & Tiemann, 2020; Lasa et al., 2020; Maass et al., 2019; Scherer & Tiemann, 2012; Shute et al., 2016). The test is a complex scenario-based mathematical problem-solving assessment within an interdisciplinary STEM context. It is designed to be practical for classroom use and to emphasize the process of solving problems. The test was proven to cover the topics and indicators for measuring mathematical problem-solving within the STEM context, although it required other skills such as science knowledge and text comprehension. Hence, the test is robust to be applied to assess mathematical problem-solving within STEM context, particularly for Indonesian students.

Mathematics DSPK has proven crucial for mathematical problem-solving (Al-Mutawah et al., 2019; Jitendra et al., 2013; Oksuz, 2009; Silao, 2018). However, existing tools primarily measure prior science knowledge or a specific topic in mathematics (Al-Mutawah et al., 2019; Binder et al., 2019; Gulacar et al., 2019; Wang et al., 2022). Additionally, the DSPK test should be a comprehensive and interconnected test to measure prior knowledge relevant to the primary assessment (i.e., an integrated STEM-based mathematical problem-solving test). Hence, a mathematics DSPK test was developed and validated.

The mathematics DSPK test measures topics that are applied in the integrated STEM-based mathematical problem-solving test, focusing on both conceptual and procedural knowledge. Hence, it serves as a comprehensive and correlated tool to measure mathematical prior knowledge in the integrated STEM-based mathematical problem-solving test. The test demonstrated consistency in measuring the mathematics DSPK, adequately covered the topics, and accurately measured indicators of procedural and conceptual mathematics knowledge. Hence, the test is robust to measure the mathematics DSPK, particularly for Indonesian students.

3.1.2 Student Profiles and Skill Development

The lack of interdisciplinary STEM tasks in mathematical problem-solving has contributed to the low achievement of Indonesian students in this area due to their unfamiliarity with the task (OECD, 2019; Suratno et al., 2020). In response to this challenge, investigating profiles and development of the skills was identified as one of the starting points to improve students' skills (Greiff et al., 2013; Molnár et al., 2013). However, issues regarding the use of monodisciplinary tasks and inconsistence results were observed in the previous studies. Therefore, the profiles and development of mathematical problem-solving skills using an integrated STEM-based mathematical problem-solving test were investigated in this study.

Interestingly, overall, Indonesian students' mathematical problem-solving skills were categorized as average, differing from the results of PISA 2018, which reported that students were ranked below average (OECD, 2019). This difference may be due to the different samples used. This study focused on high-accreditation schools (A-accreditation), while PISA 2018 used a random sample that included high-, average-, and low-accreditation schools. Although all schools in Indonesia follow the same national curriculum, high-accreditation schools are often more effective and innovative in their implementation. It impacts on stronger instructional practices, better-trained teachers, and more emphasis on higher-order thinking, including problem-solving. Therefore, these may contribute to the stronger performance observed in this sample.

Another possible factor is students' familiarity with the context of the problems used in this study. The test scenarios were based on real-life issues relevant to their local environment. It may have helped students engage more meaningfully with the tasks. In contrast, PISA items often involve general contexts that may not relate to the lived experiences of many Indonesian students. Therefore, differences in contextual familiarity may have influenced how students solved the problems (Heuvel-Panhuizen, 2005; Salgado, 2017).

Detailed results from this study revealed that Indonesian students had low performance in specific phases, such as representing and formulating, planning and executing, and monitoring and reflecting. It was also found that more students faced difficulties in the later phases of mathematical problem-solving. This aligns with research showing that students who fail to grasp mathematical concepts early tend to struggle with problem-solving later on (Novriani & Surya, 2017; Siniguian, 2017).

The best period to enhance the mathematical problem-solving skills of Indonesian students was found to be between grades 7 and 8, as their skills develop maximally. This finding is consistent with several studies in other countries (Greiff et al., 2013; Molnár et al., 2013). It shows that the development of problem-solving skills is consistent cross-culturally, considering that the measurements used are relevant for the targeted population in terms of culture, curriculum, etc. Although previous studies did not specify which subsamples should be prioritized, this study highlighted Indonesian urban school students in the knowledge acquisition phase as a focus for improvement. On the contrary, for Indonesian rural school students, the optimal development period was identified between grades 8 and 9. This may be due to the slower instructional pace in Indonesian rural schools, which might delay students' exposure to complex problem-solving tasks (OECD, 2015). Consequently, it postpones their best period of development in mathematical problem-solving skills.

3.1.3 Gender, Grade, and Regional Comparisons

Another identified solution to improve students' mathematical problem-solving skills is to examine skill differences based on background variables in order to better support underrepresented groups (Anjum, 2015; Greiff et al., 2013; Williams, 2005). However, previous studies had mixed results regarding the outperforming group, and failed to identify the specific subsamples contributing to the differences. Therefore, the differences in students' skills based on grades, school locations, and gender were explored in this study by conducting a detailed analysis of each sample's contribution.

Rural Indonesian school students performed poorly in mathematical problem-solving compared to urban school students, particularly in grades 7 and 8. These results were in agreement with studies by Ramos et al. (2021), Williams (2005), and Nepal (2017), although these studies did not investigate which students contributed to the differences. However, the findings were inconsistent with another finding that students from rural schools had better mathematical problem-solving skills (J. Lee & McIntire, 2000). J. Lee & McIntire (2000) explained that the rural schools in their study have significantly better learning conditions, such as access to instructional resources, professional development for teachers, and a supportive environment. These conditions are also typical of Indonesian urban schools. In contrast, many Indonesian rural schools face challenges such as limited infrastructure, fewer qualified teachers, fewer enrichment programs, and lower access to instructional materials (OECD, 2015). Therefore, these differences in findings may be the result of variations in facilities, policies, and human resources across contexts.

Indonesian female students outperformed male students in mathematical problem-solving skills, consistent with Anjum (2015) and Halpern et al. (2005). We found that Indonesian male students, particularly those in urban schools and in grades 7 and 9, underperformed compared to their counterparts. However, these findings were opposite to the most common findings. For instance, there were no gender differences in mathematical problem-solving due to gender equality in instructions (Ajai & Imoko, 2015; Li et al., 2018; Lindberg et al., 2010; Nepal, 2017; Vilenius-Tuohimaa et al., 2008) or males outperformed females in mathematical problem-solving (Leder, 2019).

One possible explanation for these different findings is students' attitudes toward the low-stakes problem-solving test used in this study, which may have influenced their performance. In the Indonesian context, female students may exhibit stronger motivation, greater attention, and a more serious approach to all tests, including low-stakes assessments. Additionally, school expectations and cultural norms may play a role. In many Indonesian communities, female middle school students often feel a greater pressure to perform well academically, especially in mathematics. This pressure is partly shaped by stereotypes that portray females as more diligent and responsible (Olsson et al., 2018). Therefore, female students may be more likely to invest more effort in their schoolwork, which may influence their performance on mathematical problem-solving tasks. This is the case even though instruction in Indonesian classrooms is likely conducted equally between genders.

3.1.4 Cognitive and SES Factors

Another solution proposed by previous studies to improve mathematical problem-solving skills involves examining the impact of cognitive (mathematics DSPK, science knowledge, and text comprehensive skills) and SES (parents' education and family income) factors (Akbasli et al., 2016; Al-Mutawah et al., 2019; Alghazo & Alghazo, 2015; Anjum, 2015; Boonen et al., 2014; Csapó & Molnár, 2017; Davis-Kean et al., 2021; Jitendra et al., 2013; Marks & Pokropek, 2019; Oksuz, 2009; Scherer & Beckmann, 2014; Vilenius-Tuohimaa et al., 2008). However, these studies typically analyzed the influential factors separately. Hence, these factors were examined in this study as a comprehensive model using SEM that allows for the simultaneous testing of direct and indirect relationships.

Two theoretical models were constructed, both of which obtained acceptable results. However, the model that excluded the direct path of parents' education and mathematical problem-solving exhibited better goodness-of-fit indices. Cognitive factors were the most significant predictors of Indonesian students' mathematical problem-solving skills. Strong knowledge of mathematics and science impacted students' high ability in mathematical problem-solving, which was in agreement with previous studies (Al-Mutawah et al., 2019; Bahar, 2013; Booth & Davenport, 2013; Csapó & Molnár, 2017; Jitendra et al., 2013; Oksuz, 2009; Scherer & Beckmann, 2014). Text comprehension skills also contributed directly and indirectly to Indonesian students' success in mathematical problem-solving through science knowledge. These results confirm the findings of previous studies (Akbasli et al., 2016; Anjum, 2015; Boonen et al., 2014; Salihu et al., 2018; Spencer et al., 2020; Vilenius-Tuohimaa et al., 2008). Hence, high text comprehension skills improve performance in science, further influence mathematical problem-solving skills.

Interestingly, family income was the only SES factor with a direct, even very small, effect on Indonesian students' mathematical problem-solving skills. Although this aligns with previous studies, the effect was smaller in this study compared to the standardized effects reported in earlier research (Alghazo & Alghazo, 2015; Marks & Pokropek, 2019).

Another surprising result was that mother's and father's educations did not have a direct role in influencing students' mathematical problem-solving skills. This finding contrasts with previous studies conducted in western contexts (Alghazo & Alghazo, 2015; Greiff et al., 2013; Kodippili, 2011; Pangeni, 2014). These differences may reflect cultural factors. In many western countries, education tends to be more balanced between parent-centered and teacher-centered approaches (Cortina et al., 2017; Niu et al., 2025). In contrast, education in Indonesia or a collectivist country is predominantly teacher-centered (Cortina et al., 2017; Leung, 2001; Niu et al., 2025). As a result, Indonesian students rely more on their teachers or private tutors than on their parents for academic support. Indonesian parents often place full trust in schools and tutoring centers to provide quality education, particularly in STEM subjects. Therefore, many parents focus on ensuring financial stability to afford these services and work full-time, which limits their academic interaction with their children at home. This is also because of the long hours Indonesian students typically spend at schools. Therefore, parents' education influences students' mathematical problem-solving skills indirectly, particularly through occupations and the family's financial capacity to support academic needs (Alomar, 2007; Davis-Kean et al., 2021; Hascoët et al., 2021).

This explanation is supported by our study's finding that parents' education impacted Indonesian students' mathematical problem-solving skills through family income. This result is particularly in agreement with the results of Davis-Kean et al. (2021) that parents' education plays a role in child development through family income.

Another possible reason for the lack of a direct effect is that, in Indonesia, knowledge and values are also obtained through religious education, informal education, and life experiences (OECD, 2015). Therefore, a lower level of formal education does not necessarily equate to a lack of support or understanding of their children's academic needs, even if that support is not in the form of direct mathematical instruction. Lastly, the parents' education variable in this study was based on general education levels, without considering their specific discipline or occupation. Therefore, many highly educated Indonesian parents may not have a background in mathematics or STEM. They may feel less confident in providing direct support for their children's mathematical problem-solving development.

3.1.5 Broader Reflections

This study offers more than just insights into students' ability to solve mathematical problems integrated with the STEM context, it also provides a broader view of how these skills are impacted by knowledge, background, tasks, and context. It explains the potential of interdisciplinary problem-solving assessment, particularly in the context of a developing country.

The integration of STEM contexts into mathematical problem-solving through realistic, structured, and familiar scenarios reflects a shift in how problem-solving is conceptualized. This approach invites students to engage in reasoning processes that are both culturally relevant and cognitively demanding. The findings revealed that curriculum-sensitive and interdisciplinary assessments are not only possible in mathematics but also necessary for addressing contemporary education needs.

The finding that students' mathematical problem-solving skills developed more significantly during grades 7 and 8 shows the dynamic nature of these skills. They evolve with instruction and experience. This challenges the idea that mathematical problem-solving is a fixed ability (Grissom, 2004), emphasizing instead its responsiveness to environmental factors and learning experiences. As such, this study also acknowledged that demographic location (e.g., rural vs. urban) and background, as aspects of the students' environment, play a role in students' mathematical problem-solving skills. Therefore, broader environmental and systematic factors, including quality of instruction and access to learning resources, are interdependent in shaping mathematical problem-solving skills.

The study also invites reflection on long-standing narratives about gender and mathematics. Unlike the common belief that males do better (Leder, 2019), the results showed that females outperformed males in mathematical problem-solving. Therefore, understanding how local culture, expectations, and school systems affect gender performance is important, rather than relying on general explanations.

Another critical reflection is the influence of domain-specific knowledge, particularly in mathematics and science, as a foundational prerequisite for successful problem-solving. However, text comprehension skills are also important in solving multi-step problems. Beyond cognitive domains, SES also plays a role, but its influence depends on the cultural and instructional context. In a teacher-centered educational system, like Indonesia, parents' education appears to lack direct impact. Instead, family income is more associated with students' mathematical problem-solving skills due to its effect on access to academic support and resources.

3.2 Educational Implications

Integrated STEM-based mathematical problem-solving is an obligatory skill for students according to the 2013 Indonesian mathematics curriculum (Government Regulation Number 24 Year of 2016 Attachment 15, 2016). However, low achievement and the lack of relevant studies in Indonesia have been highlighted (OECD, 2019; Suratno et al., 2020). This study contributes to the field of assessment by providing an applicable test of integrated STEM-based mathematical problem-solving. It also offers insight into students' skill development and differences among groups. Additionally, it identifies factors influencing these skills, particularly to support the improvement of Indonesian students' mathematical problem-solving skills.

There is a need for a comprehensive review of STEM assessment, specifically in problem-solving and the framework used (Gao et al., 2020). The results of the systematic review provide theoretical contributions, particularly by highlighting trends in problem-solving assessment within mathematics and STEM education. This includes general information (e.g., assessment tools, participants, and topics) and specific information (e.g., examples of the assessments and frameworks used). This information is beneficial for scholars to identify research gaps and to provide support for further studies. For example, addressing the lack of research in interdisciplinary STEM areas has been suggested as a priority.

The need for an interdisciplinary task highlights the importance for Indonesian curriculum developers to embed them and explicitly emphasize interdisciplinary competencies in the national curriculum, which are lacking in the 2013 Indonesian curriculum (Government Regulation Number 24 Year of 2016 Attachment 15, 2016). This could involve requiring a STEM project each semester and updating mathematics textbooks and assessments to reflect interdisciplinary competencies (e.g., by integrating scientific methodology into problem-solving tasks) (Kelley & Knowles, 2016; Rahimah, 2022). Additionally, it provides valuable insights for Indonesian policymakers to introduce elective interdisciplinary STEM subjects at the middle school level (OECD, 2015; Rahmaniar, 2020). This approach would allow for team-teaching or collaboration across disciplines. It also underscores the need for teacher training programs to include specific coursework on STEM integration and to assess pre-service teachers' related competencies (Rahmaniar, 2020).

The results of the review also guide global and Indonesian teachers and researchers in deciding on the appropriate tools to be adapted or a new assessment tool to be developed based on the frameworks, challenges, drawbacks, and advantages addressed in this study. For example, (1) a suggestion to use a non-routine essay test to measure a cognitive monodisciplinary area; (2) consideration of psychometric evidence and cultural issues in interdisciplinary STEM tools; (3) a recommendation to incorporate metacognitive prompts in interdisciplinary STEM tests, etc.

The integrated STEM-based mathematical problem-solving test contributes to addressing several challenges: the demand for interdisciplinary STEM tasks, the underrepresentation of mathematics, the complexity and product-based of existing tests, and their limited applicability in classroom (Annaggar & Tiemann, 2020; Lasa et al., 2020; Maass et al., 2019; Scherer & Tiemann, 2012; Shute et al., 2016). This test offers a classroom-friendly and process-based assessment for Indonesian teachers. It highlights the interdisciplinary STEM context through specific scientific phenomena in Indonesia. It is designed to be practical and time-efficient due to the flexibility of using only one scenario to assess students' mathematical problem-solving skills. This is possible because each scenario covers all necessary steps in mathematical problem-solving. The novel use of metacognitive prompt items ensures that students' responses are consistent, accurately addresses students' abilities, and explores all problem-solving indicators. This approach helps address the gap left by product-based assessment in STEM problem-solving. Additionally, the test has had several validity and reliability tests and has been shown to be appropriate for use by Indonesian middle school students. Hence, the test addresses not only the need for a 21st century challenge but also the needs of curriculum, students, and teachers in Indonesia. In general use, the test can be adapted to other contexts with attention to cross-cultural considerations. Psychometric evidence of the test can be a basis for other relevant studies.

The integrated STEM-based mathematical problem-solving test can also be used as a model tool for practicum or training modules in teacher training programs. It helps pre-service teachers learn how to implement and analyze interdisciplinary process-based problem-solving assessments. Additionally, it provides policymakers with a robust example of an assessment tool that aligns with the Indonesian curriculum and addresses current educational needs. Furthermore, the test can be recommended for large-scale implementation as a diagnostic tool to inform the development of an Indonesian national STEM curriculum or to revise the integration of interdisciplinary elements into the existing mathematics curriculum (Rahmaniar, 2020).

The mathematics DSPK test contributes to solving issues in existing instruments, such as their science-oriented, focus on a single mathematics topic, and the unavailability of a comprehensive and correlated test to measure prior knowledge in the integrated STEM-based mathematical problem-solving test (Al-Mutawah et al., 2019; Binder et al., 2019; Gulacar et al., 2019; Wang et al., 2022). The test contributes to assessment practices in Indonesia by providing a validated test to measure prior mathematical knowledge, both procedural and conceptual, in some mathematical topics. The test can be used for Indonesian practitioners as a preliminary assessment for mathematical problem-solving and aligns with the Indonesian curriculum and context, while also infusing elements of international curricula. Hence, it provides a comprehensive assessment tool that supports both assessment of mathematical problem-solving in the STEM context and prior mathematical knowledge. Globally, this test is also possible to be adapted for different populations. The results of the test validity can be a basis for adapting the test to a different sample background and for further testing of its validity.

The findings on students' mathematical problem-solving skills help teachers to emphasize teaching and learning in a specific problem-solving phase where students struggle the most, such as in representing and formulating, planning and executing, and monitoring and reflecting phases. For Indonesian curriculum developers, it is recommended to address the gaps in the 2013 Indonesian mathematics curriculum by: (1) explicitly defining each step in the problem-solving process in the national mathematics curriculum guidelines; (2) including tasks and examples in the mathematics textbooks that align with these phases; and (3) revising both summative and formative assessments to measure each process step (Government Regulation Number 24 Year of 2016 Attachment 15, 2016; Rahimah, 2022; Suurtamm et al., 2016). For Indonesian policymakers, it is advisable to incorporate process-based mathematical problem-solving as a core competency in national education policy and allocate resources for related professional development (Chang et al., 2013; OECD, 2015). Additionally, teacher training programs should include modules with practical exercises using a phase-based problem-solving task to diagnose students' difficulties across phases.

Detailed findings on development and differences in these skills help address the issue related to which subsamples contribute to the differences by implementing a STEM-based task (Anjum, 2015; Greiff et al., 2013; Molnár et al., 2013; Williams, 2005). The results also guide teachers in designing appropriate instructional strategies, particularly for grades 7 and 8, such as mathematics experimentation, inquiry-based learning, and engineering-based design activities (Kelley & Knowles, 2016; Molnár et al., 2013). Policymakers can support this by facilitating training programs for grades 7 and 8 mathematics teachers in designing appropriate mathematical problem-solving instructions. Additionally, the findings can inform curriculum and policy reforms that emphasize hands-on, strategy-based problem-solving activities between grades 7 and 8. This is particularly important for urban school students during the knowledge acquisition phase, when their skills grow maximally.

The findings also provide theoretical implications for ongoing debates on skill differences among student groups (Ajai & Imoko, 2015; Anjum, 2015; Leder, 2019; J. Lee & McIntire, 2000; Li et al., 2018; Nepal, 2017; Ramos et al., 2021; Vilenius-Tuohimaa et al., 2008). For practical implications, the descriptions of the differences guide educational policymakers, particularly in Indonesia, to build an equitable educational system for underrepresented groups. For instance, to reduce regional disparities, targeted reforms may include (1) increased funding and resource allocation for rural schools, particularly for grades 7 and 8; (2) standardized criteria for mathematics teacher recruitment in rural schools by aligning with those used in urban schools; and (3) urban-rural school partnerships to promote exchange innovations and best practices (Muttaqin, 2017; OECD, 2015). Additionally, the results also encourage rural school teachers to adopt effective strategies used in urban settings for teaching mathematical problem-solving, while adapting them to local needs. This can be supported through professional development programs designed by curriculum developers and facilitated by policymakers.

Given the underrepresentation of male students in mathematical problem-solving, curriculum contents and instructions should explicitly address gender stereotypes by using examples and tasks that are suitable to diverse students' interest. To support this, policymakers should mandate gender-focused teacher training programs.

The developed model of factors influencing mathematical problem-solving skills can be replicated for relevant studies in this area, as it demonstrated statistical robustness and suitability for the analyzed data. Additionally, the results of direct and indirect effects affecting these skills provide theoretical implications, addressing gaps in studies that fail to consider all relevant factors simultaneously (Akbasli et al., 2016; Al-Mutawah et al., 2019; Alghazo & Alghazo, 2015; Anjum, 2015; Boonen et al., 2014; Davis-Kean et al., 2021; Jitendra et al., 2013; Marks & Pokropek, 2019; Oksuz, 2009; Scherer & Beckmann, 2014; Vilenius-Tuohimaa et al., 2008). For practical implications, these results guide teachers to design instructions by emphasizing cognitive factors, specifically for the Indonesian population. It can be through recalling and improving mathematics and science concepts when such knowledge is needed but appears to be poor. Familiarizing students with problem-solving tasks can also improve text comprehension skills, which enhances science performance and further influences mathematical problem-solving skills. These cognitively focused strategies should be included in teacher training modules. For Indonesian curriculum developers, allocating specific time in the curriculum and lesson plans to review prior knowledge before teaching mathematical problem-solving is recommended (Government Regulation Number 24 Year of 2016 Attachment 15, 2016). Mathematics textbooks and teacher guides should also highlight essential prerequisite concepts and provide diagnostic tools to assess students' readiness (Rahimah, 2022). Lastly, policymakers should provide support and resources for economically disadvantaged students that can also impact their mathematical problem-solving skills (OECD, 2015).

3.3 Limitations and Future Directions

The general limitations of this study include its cross-sectional study design and the use of a sample from only A-accreditation schools in East Java, Indonesia. Therefore, the results may reflect a high performing educational setting and do not represent the diversity of students in lower-accredited schools or other regions. Additionally, the cross-sectional study is limited to determine the long-term stability of the results of the mathematical problem-solving skills of the students and the factors influencing them. Therefore, a longitudinal study with a more varied sample is necessary to produce broader generalization.

Another limitation is that the participants were from consecutive grades, which might have caused some differences and the development of the mathematical problem-solving skills in specific subsamples were not significant because the level of their skills was similar. Hence, it is suggested for future studies to expand the participants from upper elementary, middle school, and lower high-school to capture a wider range of abilities and developmental stages.

The PRISMA framework was used to review problem-solving assessment in STEM education. However, methodological limitations include the small number of reviewers, the overly broad inclusion criteria, and the inclusion of grey literature. These limitations may impact the consistency and rigor of the review process and led to a less specific conclusion of the results. Hence, it is suggested to include at least three reviewers and consider only articles from high-impact factor journals for further study to enhance research quality. Additionally, the inclusion criteria should be more precise (e.g., focusing only on non-experimental study) and provide detailed psychometric evidence of the reviewed articles.

The paper-based format and the lack of technology and engineering role in the content used in the integrated STEM-based mathematical problem-solving test are acknowledged as limitations of this test. It is suggested to develop a computer version of the test and train participants on the related technology used before data collection. Additionally, the integration of engineering and technology should be emphasized through the inclusion of digital or hands-on tasks in the test. Another limitation is that the mathematics DSPK test is a multiple-choice format, which allows the possibility of guessing. This may have contributed to the unacceptable behavior of the distractors in several items. Therefore, the mathematics DSPK test should be complemented with open-ended tasks. Additionally, the absence of researchers in assessing the content validity of both integrated STEM-based mathematical problem-solving and mathematics DSPK tests encourages future studies to involve both practitioners and researchers in this process.

The limited number of background variables and factors that influence mathematical problem-solving skills narrows the results. The conclusions from the study may not fully capture the complexity of the factors influencing these skills. Therefore, it is encouraged for future studies to add more potential variables, such as working memory and parental job. Additionally, the use of a broad instrument to gather information on the SES of students may impact the non-significant findings or weak effects of those variables in influencing mathematical problem-solving skills. Therefore, it is also suggested to use a more precise questionnaire to assess parents' education by focusing on their discipline and the depth of their mathematical knowledge. Hence, the investigation of the influence of parents' education can be more in-depth.

3.4 References

Ajai, J. T., & Imoko, B. I. (2015). Gender differences in mathematics achievement and retention scores: A case of problem-based learning method. *International Journal of Research in Education and Science*, *1*(1), 45–50. https://doi.org/10.21890/ijres.76785

Akbasli, S., Sahin, M., & Yaykiran, Z. (2016). The effect of reading comprehension on the performance in science and mathematics. *Journal of Education and Practice*, *7*(16), 108–121. https://files.eric.ed.gov/fulltext/EJ1108657.pdf

Al-Mutawah, M. A., Thomas, R., Eid, A., Mahmoud, E. Y., & Fateel, M. J. (2019). Conceptual understanding, procedural knowledge and problem-solving skills in mathematics: High school graduates work analysis and standpoints. *International Journal of Education and Practice*, *7*(3), 258–273. https://doi.org/10.18488/journal.61.2019.73.258.273

Alghazo, Y., & Alghazo, R. (2015). The effect of parental involvement and socioeconomic status on elementary students' mathematics achievement. *Journal of Social Sciences and Humanities*, *1*(5), 521–527. http://www.aiscience.org/journal/paperInfo/jssh?paperId=2312

Alomar, B. O. (2007). Personal and family factors as predictors of pupils' mathematics achievement. *Psychological Reports*, *101*(1), 259–269. https://doi.org/10.2466/PR0.101.1.259-269

Anjum, S. (2015). Gender difference in mathematics achievement and its relation with reading comprehension of children at upper primary stage. *Journal of Education and Practice*, *6*(16), 71–75. https://files.eric.ed.gov/fulltext/EJ1079951.pdf

Annaggar, A., & Tiemann, R. (2020). A video game as an assessment tool of problem-solving competence. *Gesellschaft Fur Didactik Der Chemie Und Physik (GDCP)*, *April*, 377–380. https://www.gdcp-ev.de/wp-content/tb2020/TB2020_377_Annaggar.pdf

Bahar, A. (2013). *The influence of cognitive abilities on mathematical problem solving performance* [The University of Arizona]. https://repository.arizona.edu/handle/10150/293594

Binder, T., Sandmann, A., Sures, B., Friege, G., Theyssen, H., & Schmiemann, P. (2019). Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression. *International Journal of STEM Education*, *6*(1). https://doi.org/10.1186/s40594-019-0189-9

Boonen, A. J. H., Van Wesel, F., Jolles, J., & Van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research*, *68*, 15–26. https://doi.org/10.1016/j.ijer.2014.08.001

Booth, J. L., & Davenport, J. L. (2013). The role of problem representation and feature knowledge in algebraic equation-solving. *Journal of Mathematical Behavior*, *32*(3), 415–423. https://doi.org/10.1016/j.jmathb.2013.04.003

Chang, M. C., Shaeffer, S., Al-Samarrai, S., Ragatz, A. B., de Ree, J., & Stevenson, R. (2013). Teacher Reform in Indonesia:The Role of Politics and Evidence in Policy Making. In *Teacher Reform in Indonesia:The Role of Politics and Evidence in Policy Making*. https://doi.org/10.1596/9780821398296

Cortina, K. S., Arel, S., & Smith-Darden, J. P. (2017). School Belonging in Different Cultures: The Effects of Individualism and Power Distance. *Frontiers in Education*, *2*(November), 1–12. https://doi.org/10.3389/feduc.2017.00056

Csapó, B., & Molnár, G. (2017). Potential for assessing dynamic problem-solving at the beginning of higher education studies. *Frontiers in Psychology*, *8*(NOV). https://doi.org/10.3389/fpsyg.2017.02022

Davis-Kean, P. E., Tighe, L. A., & Waters, N. E. (2021). The role of parent educational attainment in parenting and children's development. *Current Directions in Psychological Science*, *30*(2), 186–192. https://doi.org/10.1177/0963721421993116

Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education*, *7*(1). https://doi.org/10.1186/s40594-020-00225-4

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts-something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, *105*(2), 364–379. https://doi.org/10.1037/a0031856

Grissom, J. B. (2004). Age and achievement. Educational Policy Analysis Archives, 12(49). https://doi.org/https://doi.org/10.14507/epaa.v12n49.2004
Gulacar, O., Milkey, A., & Mclane, S. (2019). Exploring the effect of prior knowledge and gender on undergraduate students' knowledge structures in chemistry. *Eurasia Journal of Mathematics, Science and Technology Education*, *15*(8). https://doi.org/10.29333/ejmste/106231
Halpern, D. F., Wai, J., & Saw, A. (2005). A psychobiosocial model: Why females are sometimes greater than and sometimes less than males in math achievement. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 48–72). Cambridge University Press.
Hascoët, M., Giaconi, V., & Jamain, L. (2021). Family socioeconomic status and parental expectations affect mathematics achievement in a national sample of Chilean students. *International Journal of Behavioral Development*, *45*(2), 122–132. https://doi.org/10.1177/0165025420965731
Heuvel-Panhuizen, M. Van Den. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics*, *25*(2), 2–9. https://www.jstor.org/stable/40248489
Government Regulation Number 24 Year of 2016 Attachment 15, 1 (2016).
Jitendra, A. K., Lein, A. E., Star, J. R., & Dupuis, D. N. (2013). The contribution of domain-specific knowledge in predicting students' proportional word problem-solving performance. *Educational Research and Evaluation*, *19*(8), 700–716. https://doi.org/10.1080/13803611.2013.845107
Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1). https://doi.org/10.1186/s40594-016-0046-z
Kodippili, A. (2011). Parents' education level in students' mathematics achievement; do school factors matter? *Academic Leadership*, *9*(1). https://scholars.fhsu.edu/alj/vol9/iss1/39/
Lasa, A., Abaurrea, J., & Iribas, H. (2020). Mathematical content on STEM activities. *Journal on Mathematics Education*, *11*(3), 333–346. https://doi.org/10.22342/JME.11.3.11327.333-346
Leder, G. C. (2019). Gender and mathematics education: Overview. In G. Kaiser & N. Presmeg (Eds.), *Compendium for early career researchers in mathematics education* (pp. 289–308). Springer Open. https://doi.org/10.1007/978-3-030-15636-7_7
Lee, J., & McIntire, W. G. (2000). Interstate variation in the mathematics achievement of rural and nonrural students. *Journal of Research in Rural Education*, *16*(3), 168–181. https://jrre.psu.edu/sites/default/files/2019-08/16-3_2.pdf
Leung, F. K. S. (2001). In Search of an East Asian Identity in Mathematics Education. *Educational Studies in Mathematics*, *47*, 35–51. https://doi.org/10.1007/978-94-010-9046-9_34
Li, M., Zhang, Y., Liu, H., & Hao, Y. (2018). Gender differences in mathematics achievement in Beijing: A meta-analysis. *British Journal of Educational Psychology*, *88*(4), 566–583. https://doi.org/10.1111/bjep.12203
Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and

mathematics performance: A meta analysis. *Psychology Bulletin*, *136*(6), 1123–1135. https://doi.org/10.1037/a0021276.New

Maass, K., Geiger, V., Ariza, M. R., & Goos, M. (2019). The role of mathematics in interdisciplinary STEM education. *ZDM - Mathematics Education*, *51*(6), 869–884. https://doi.org/10.1007/s11858-019-01100-5

Marks, G. N., & Pokropek, A. (2019). Family income effects on mathematics achievement: their relative magnitude and causal pathways. *Oxford Review of Education*, *45*(6), 769–785. https://doi.org/10.1080/03054985.2019.1620717

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, *9*, 35–45. https://doi.org/10.1016/j.tsc.2013.03.002

Muttaqin, T. (2017). The education divide in Indonesia: Four essays on determinants of unequal access to and quality of education. In *Thesis fully internal, University of Groningen*. University of Groningen. https://research.rug.nl/files/40126192/Summary_in_English.pdf

Nepal, B. (2017). Impact of gender and location on mathematical thinking and mathematics achievement. *Journal of Advanced Academic Research*, *3*(3), 11–21. https://doi.org/10.3126/jaar.v3i3.16803

Niu, J., Xu, H., & Yu, J. (2025). Identifying multilevel factors on student mathematics performance for Singapore, Korea, Finland, and Denmark in PISA 2022: considering individualistic versus collectivistic cultures. *Humanities and Social Sciences Communications*, *12*(1), 1–12. https://doi.org/10.1057/s41599-025-04466-y

Novriani, M. R., & Surya, E. (2017). Analysis of Student difficulties in mathematics problem solving ability at MTS Swasta IRA Medan. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, *33*(3), 63–75. https://core.ac.uk/download/pdf/249335829.pdf

OECD. (2015). Education in Indonesia: Rising to the Challenge. In *OECD*. OECD Publishing. https://doi.org/10.2307/3023860

OECD. (2019). Programme for International Student Assessment (PISA). Results From PISA 2018: Indonesia. *OECD Education Working Papers*. https://doi.org/10.1007/978-94-6209-497-0_69

Oksuz, C. (2009). Association of domain-specific knowledge and analytical ability with insight problem solving in mathematics. *International Journal of Pedagogies and Learning*, *5*(1), 138–153. https://doi.org/10.5172/ijpl.5.1.138

Olsson, J., Olsson, S., & Jobér, A. (2018). *Gender within an Indonesian Mathematics Classroom* (Issue March). Malmo University. https://www.diva-portal.org/smash/get/diva2:1491288/FULLTEXT01.pdf

Pangeni, K. P. (2014). Factors determining educational quality: Student mathematics achievement in Nepal. *International Journal of Educational Development*, *34*(1), 30–41. https://doi.org/10.1016/j.ijedudev.2013.03.001

Rahimah, D. (2022). *Mathematics Textbooks as Curriculum Resources: Exploring How the Ministry-Published Textbook Supports Indonesian Teachers in Implementing a Student-Centred Teaching Approach*. *October*, 169. https://espace.library.uq.edu.au/view/UQ:0bb4707/s4428543_phd_thesis.pdf

Rahmaniar, A. (2020). *Perceptions of the Indonesia National Curriculum in Relation To Integrated Stem Education At the High School Level*. 1–99. https://ir.library.illinoisstate.edu/cgi/viewcontent.cgi?article=2341&context=etd

Ramos, R., Duque, J. C. C., & Nieto, S. (2021). Decomposing the rural-urban differential in student achievement in Colombia using PISA microdata. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2051358

Salgado, A. J. F. (2017). The role of context and context familiarity on mathematics problems. *Revista Latinoamericana de Investigacion En Matematica Educativa*, *20*(3), 265–292. https://doi.org/10.12802/relime.17.2031

Salihu, L., Aro, M., & Räsänen, P. (2018). Children with learning difficulties in mathematics: Relating mathematics skills and reading comprehension. *Issues in Educational Research*, *28*(4), 1024–1038. https://www.semanticscholar.org/paper/Children-with-learning-difficulties-in-mathematics%3A-Salihu-Aro/5a9c7f13502ebc431d07f769331763030cfdcd4e

Scherer, R., & Beckmann, J. F. (2014). The acquisition of problem solving competence: evidence from 41 countries that math and science education matters. *Large-Scale Assessments in Education*, *2*(1), 1–22. https://doi.org/10.1186/s40536-014-0010-7

Scherer, R., & Tiemann, R. (2012). Factors of problem-solving competency in a virtual chemistry environment: The role of metacognitive knowledge about strategies. *Computers and Education*, *59*(4), 1199–1214. https://doi.org/10.1016/j.compedu.2012.05.020

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106–117. https://doi.org/10.1016/j.chb.2016.05.047

Silao, I. V. (2018). Factors affecting the mathematics problem solving skills of Filipino pupils. *International Journal of Scientific and Research Publications*, *8*(2), 487. https://doi.org/https://www.ijsrp.org/research-paper-0218/ijsrp-p7461.pdf

Siniguian, M. T. (2017). Students difficulty in solving mathematical problems. *International Journal of Advanced Research in Engineering and Applied Sciences*, *6*(2), 1–12. https://garph.co.uk/IJAREAS/Feb2017/1.pdf

Spencer, M., Fuchs, L. S., & Fuchs, D. (2020). Language-related longitudinal predictors of arithmetic word problem solving: A structural equation modeling approach. *Contemporary Educational Psychology*, *60*(November 2019), 101825. https://doi.org/10.1016/j.cedpsych.2019.101825

Suratno, Wahono, B., Chang, C. Y., Retnowati, A., & Yushardi, Y. (2020). Exploring a direct relationship between students' problem-solving abilities and academic achievement: A STEM education at a coffee plantation area. *Journal of Turkish Science Education*, *17*(2), 211–224. https://doi.org/10.36681/tused.2020.22

Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S., & Vos, P. (2016). Assessment in Mathematics Education. In *Debates in Mathematics Education*. https://doi.org/10.4324/9780429021015-19

Vilenius-Tuohimaa, P. M., Aunola, K., & Nurmi, J. E. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology*, *28*(4), 409–426. https://doi.org/10.1080/01443410701708228

Wang, J., Stebbins, A., & Ferdig, R. E. (2022). Examining the effects of students' self-efficacy and prior knowledge on learning and visual behavior in a physics game. *Computers and Education*, *178*(May 2021), 104405. https://doi.org/10.1016/j.compedu.2021.104405

Williams, J. H. (2005). Cross-national variations in rural mathematics achievement: A descriptive overview. *Journal of Research in Rural Education*, *20*(5), 1–18. https://jrre.psu.edu/sites/default/files/2019-08/20-5.pdf

APPENDIX A

AN INTEGRATED STEM-BASED MATHEMATICAL PROBLEM-SOLVING TEST (English Version)

The Construct of the Test

Indicators	Total of items	Items
Determining the goal of a problem	3 items in every grade	Item number 1 in every scenario and in every grade
• Identifying relevant and valuable unknown and given information • Representing the problem by constructing tabular, graphical, symbolic, or verbal representations using technology	3 items in every grade	Item number 2 in every scenario and in every grade
Employing a useful basic concept	3 items in every grade	Item number 3 in every scenario and in every grade
Making a prediction (hypothesis), developing criteria, or checking existing theory	3 items in every grade	Item number 4 in every scenario and in every grade
Formulating a reasonable argument, explanation, and solution (including strategy, step design, and model building)	3 items in every grade	Item number 5 in every scenario and in every grade
Arranging, critically choosing, and evaluating alternatives by deducting, proofing, finding a counterexample	3 items in every grade	Item number 6 in every scenario and in every grade
Performing a plan by applying mathematics concepts, mathematization, reasoning, computational skills, science concept, and technology	3 items in every grade	Item number 7 in every scenario and in every grade
Drawing conclusions, evaluating, and reflecting on results and methods	3 items in every grade	Item number 8 in every scenario and in every grade

The Instrument

Scenario: Eco-Friendly Packaging

Eco-friendly and low-budget product is starting to be prioritized. We want to produce a packaging with the criteria:

1. The packaging has length, width, and height of 18cm x 18cm x 8cm respectively
2. The cover of the packaging is square with 1.2 times the area of the base.
3. There are four material options:

	Duplex paper	Ivory paper	Styrofoam	Plastic
Size	79cm×109cm	$215mm \times 330mm$	100cm × 50cm	17cm×50cm
Thickness	250gsm	250gsm	0.5cm	0.5cm
Price	Rp 4,000.00	Rp 2,000.00	Rp 9,000.00	Rp 1,300.00
The time needed to decompose	2-6 months		Can't decompose	1000 years

Challenge

Design an eco-friendly packaging and decide the lowest budget of a packaging production based on these criteria!

Prompting questions:

1. What is your challenge (asked) in the scenario?
2. What information do you need to solve the challenge?
3. What is the total area of four sides, cover, and base in cm^2?
4. What is your guess relating to the type of material that could be used to design the packaging? Give your reasons.
5. Draw your packaging design based on the shape and length.
6. Mention two materials that are unsuitable to be chosen, and give your reasons!
7. How much is the price for your packaging based on your chosen material and area of your packaging?
8. Is the packaging eco-friendly and the cheapest one? Give your conclusion about the price and type of material used to produce packaging.

Scenario: School Park

Planting trees is a program for protecting the Earth from the Ministry of Education. A "Tunas Bangsa" school has a park for planting with an area of $20m^2$. The ratio between width and length is 1:5. Students have to plant a type of tree around the park in December 2020. There are some options of trees that could be planted with an ideal distance in every tree is 0.5m. The effectiveness of a tree to absorb CO2 will increase $\frac{1}{2}$ times from the previous year

	Absorption of CO2 (kg/tree/year)	Diameter (cm)
Annona muricata	75.29	30
Mimusops elengi	34.29	20
Khaya anthotheca	21.29	120

Challenge

Design a school park to plant one of the most effective type of tree that can absorb CO2 emissions and draw a graph that represents the total of CO2 that could be absorbed by the planted trees in 2021, 2022, and 2023.

Prompting questions:

1. What is your challenge (asked) in the scenario?
2. What information do you need to solve the challenge?
3. The trees only can be planted around the park. Based on the information, choose the concept of circumference or area that will be applied. What is the area/ circumference of the park (based on your choice) in m or m^2?
4. What is your guess relating to the type of tree that could be planted, and give the reason!

5. Draw the park design based on the size, the position of trees, the number of trees, the diameter of trees, and the distance between each tree!
6. Mention a type of tree that is unsuitable to be chosen, and give your reason!
7. Use your grid paper to draw a graph representing the total CO2 that can be absorbed by the planted trees in 2021, 2022, and 2023. Please use the consistency scale to draw the graph.
8. Is the tree the most effective type of tree to be planted? What is the total amount (in kg) of CO2 absorbed by trees in 2021, 2022, and 2023 respectively?

Scenario: Calory Versus Greenhouse Gas Emission

Food produces CO2 emissions in the production process, which sometimes is not balanced with the calory contained in that food. The average calorie needed for a man from their main menu is 1725 kcal and for a woman is 1500 kcal per day. The table below shows information about nutrition facts and CO2 emission in food production.

Food	Calory (in kcal)	Protein (in kcal)	Fat (in kcal)	Carbohydrate (in kcal)	CO2 emission in food production (in kg)
Salad	5	2.52	7.2	51.6	1.03
Pepes tahu (tofu with spicy seasoning) + rice	638	56.4	39.6	362.48	2.43
Beef cheese burger with oven-fried potato	649	63.2	169.2	360	25.82
Fried rice with omelet	455	104	249.3	290.4	2.29
Rendang (beef with Indonesia seasoning) + rice	841	230	256.5	421.2	25.76
Sate kambing with lontong (grilled lamb with rice cake)	548	86.8	128.7	162.4	4.46

Challenge

Design your a-day food menu based on the provided list menu. The requirements are:

1. Choose different three foods that fulfill more than the average calory needed
2. Fulfill more than 25%, 60%, and 15% of the average calory needed for fat, carbohydrate, and protein, respectively
3. After fulfilling requirements 1 and 2, choose the foods with the least CO2 emission.

Prompting questions:

1. What is your challenge (asked) in the scenario?
2. What information do you need to solve the challenge?
3. What is the minimum of fat, carbohydrate, and protein needed to fulfill the criteria (in kcal)?
4. What is your guess relating to 3 foods to be chosen as your a-day menu?
5. What is your reason relating to your choice in number 4?
6. Mention 2 foods that are unsuitable to be chosen, and give your reasons!
7. Calculate each of the calories, protein, fat, and carbohydrate from three of your chosen foods (in kcal), respectively.
8. Give the conclusion by mentioning three foods that you chose. Are they fulfilling the criteria of calory, fat, protein, and carbohydrate?

Scenario: Flood Water Reservoir

The government wants to build a rectangular prism flood water reservoir with pumps inside the water reservoir in underpass “Mayjen”. The information is:

1. The volume of the water reservoir is decided from the average volume of rainwater. The average volume of rainwater is determined by the average rainwater intensity. The data

shows the rainwater intensity in the highest rainfall month, that is November (in the mm/hour) 17, 22, 15, 17, 15, 22, 21, 11, 9, 17, 11, 8, 8, 9, 10, 17, 21, 11, 9, 10
Note: in this case, 1mm/hour equals 50liter

2. There are several options for pumps package to suck the total volume of rainwater in the water reservoir. Choose an effective package to be put in the water reservoir.

	The number of pumps	The amount of water that could be sucked per hour/ pump
Package 1	4	30liter
Package 2	3	25liter
Package 3	2	70liter

Challenge
Design a water reservoir (and its pump) and decide the fastest time for your chosen pump package to suck the average water

Prompting questions:
1. What is your challenge (asked) in the scenario?
2. What information do you need to solve the challenge?
3. What is the average volume of rainwater (in liters) in November? Use one of the suitable data representations (table/graph) to represent the rainwater intensity data.
4. What is your guess relating to the most suitable pump package to be put in the water reservoir, and give your reason!
5. Design the water reservoir (size and shape) with your chosen pump package?
6. Mention a package that is unsuitable to be selected in the water reservoir. Give your reason!
7. How many hours are the time needed for pumps in a package to suck the total volume in the water reservoir without remaining based on your guessing package?
8. Is your chosen package the most effective pump package to suck the water? Give the conclusion, how many hours are the time needed for pumps in a package to suck the average volume in the water reservoir without remaining?

Scenario: City Park
SIER is one of the most polluted areas in Surabaya. The average CO2 emission from transportation is predicted to become 4 million kg, 4.2 million kg, and 4.4 million kg in 2021, 2022, and 2023 respectively. We will plant a type of tree in a rectangular area of SIER park with an area of $21m^2$. The ratio of its sides is 7:3. There are some options of trees that could be planted with an ideal distance in every tree is 50cm.

	Absorption of CO2 (kg/tree/year)	Diameter
Samanea saman	28.448,39	15 meters
Cassia sp	5.295,47	1 meter
Canangium podratum	756,59	1 meter

Challenge
Design a rectangular area in the SIER park to plant one of the most effective types of tree that can absorb CO2 emissions and draw the graph of the remaining amount of CO2 emission from transportation in 2021, 2022, and 2023 after planting the trees.

Prompting questions:
1. What is your challenge (asked) in the question?
2. What information do you need to solve the challenge?
3. How many meters are the width and length of a planted area in a park?
4. What is your guess relating to the suitable type of tree to be planted in a park in SIER, and give your reason!
5. Draw a park that contains the shape and size, position of trees, the number of trees, and the distance between trees!

6. Mention two types of trees that are unsuitable to be planted in the park based on the criteria, and give your reasons!
7. Use your grid paper to draw a graph of the remaining CO2 emission from transportation in 2021, 2022, and 2023 after planting the trees in a planted greenery area! Use the consistency scale to draw a graph.
8. Is the type of tree the most suitable tree to be chosen? Give the conclusion related to how many kg is the remaining CO2 emission after planting the trees in 2021, 2022, and 2023 respectively.

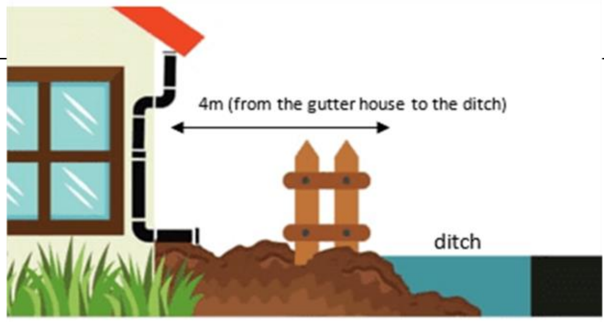
Scenario: Infiltration Well

One of the methods to increase the availability of clean water is building an infiltration well. You have a project to build an infiltration well in Mr. Adi's house. The criteria are:

1. The well has a distance of 1m from the gutter house.
2. You need to add 0.4m for turning a pipe into underground (the picture showed that the pipe ends on the ground)
3. It has a diameter of 100cm and a height of 1.5m
4. To strengthen the wall of the well, we use brick (a side area of a brick= 10 cm x 22cm, for every 40 bricks we need 11.5kg sand, and the ratio of sand and cement is 4:1)
5. There is a water inlet that drains rainwater from gutters into infiltration wells using paralon pipes
6. There is drainage from the infiltration well to the ditch that removes excess water when the infiltration well has excess water. The height of the drain pipe must be higher than the highest groundwater level in the road ditch.
7. The price list of the materials needed for building the infiltration well is mentioned as follows

Materials	Price
Brick	Rp 725/block
Paralon pipe	Rp 24.000/meter
Cement	Rp 61.000/pcs (40kg)
Sand	Rp 3200/10 kg

Mr. Adi's House



Challenge

Design an infiltration well and calculate the price needed to buy materials to build the well

Prompting questions:

1. What is your challenge (asked) in the scenario?
2. What information do you need to solve the challenge?
3. How many total bricks, total pipe (in meters), total cement (in kg), and sand (kg) that you need to build the well? Note: you need to use 10cm as the height of the brick and 22cm as the length of the brick.
4. Make a guess relating to the position of the infiltration well!
5. Design the infiltration well based on distance, shape, size from the given specifications
6. How many meters are the position of the infiltration well from the ditch and the gutter, respectively?
7. How much is the material price for building the infiltration well based on your design?
8. Are you buy more pipe, cement, and sand than you need? Give your conclusion about the total material price for building the infiltration well!

What is your gender?	Boy Girl
What grade are you in?	Grade 7 Grade 8 Grade 9
How old are you?	11 years old 12 years old 13 years old 14 years old 15 years old 16 years old
In which type of school do you study?	Public school Private School International School
Where is your school located?	Surabaya Sidoarjo Other……
What is your ethnicity?	Javanese Madurese Sundanese Other…..
What is your mother (and father) last educational level?*	Primary School Middle School High School Bachelor Master Doctorate
What is your father (and mother) job?*	Doctor Police Techer Lecturer Engineer Other….
What is the range average of your father's (and mother's) salary?*	Less than IDR 1 million IDR 1 million to IDR 2,999,000 IDR 3 million to IDR 5 million More than 5 million
How many siblings do you have?	

*the question is divided into two different items

The Construct of the Scoring Rubric

Score	Descriptions
5 (a complete and correct answer)	• Identify all relevant and valuable unknown and given information • Determine the goal completely and relevant to the problem • Employ all the basic concepts correctly with appropriate and correct calculation • Represent the problem by constructing tabular, graphical, symbolic, or verbal representations using technology completely, correctly, and appropriately. • Make a complete prediction (hypothesis), develop complete criteria, or check existing theory correctly and logically connected to the situation • Formulate the reasonable argument, explanation, and solution (including strategy, steps design, and model building) completely and correctly. • Arrange, critically choose, and evaluate alternatives by deducting, proofing, finding a counterexample appropriately and correctly. • Perform a plan by applying mathematics concepts, mathematization, reasoning, computational skills, and technology correctly, appropriately, and completely • Draw logical conclusions, evaluation, and reflection on results and methods completely and correctly
4 (a complete answer with minor error)	• Identify all relevant and valuable unknown and given information with minor errors • Determine the goal completely and relevant to the problem with minor errors • Employ all the basic concepts with minor calculation errors • Represent the problem completely but contain minor errors • Make a complete prediction (hypothesis), develop complete criteria, or check existing theory logically connected to the situation but contain minor errors • Formulate the reasonable argument, explanation, and solution (including strategy, steps design, and model building) completely but contain a minor error • Arrange, critically choose, and evaluate alternatives with minor errors • Perform a plan completely with minor errors • Make logical conclusions, evaluate, and reflect on results and methods completely with minor errors
3 (an incomplete answer but correct OR complete answer with major error)	• Identify any of the relevant and valuable unknown or given information • Determine an incomplete goal relevant to the problem OR determine the goal completely with major errors • Employ any of the basic concepts correctly with appropriate and correct calculation OR employ all the basic concepts with major errors • Represent the problem appropriately and correctly but incomplete OR represent the problem completely with major errors • Make prediction (hypothesis), develop criteria, or check an existing theory connected to the situation incompletely without error OR make a complete prediction (hypothesis), develop complete criteria, or check existing theory logically connected to the situation but contain major errors • Formulate an incomplete of the reasonable argument, explanation, and solution (including strategy, steps design, and model building) correctly OR give complete argument, explanation, and solution with major errors • Arrange, critically chose, and evaluate alternatives with major errors • Perform the plan correctly appropriately, but incompletely OR carry out the plan completely with major errors • Make incomplete conclusions, evaluation, and reflection on results and methods correctly OR make complete conclusion, evaluating, and reflecting with major errors

The Construct of the Scoring Rubric (cont.)

Score	Descriptions
2 (an incomplete and contain minor error)	• Identify any of the relevant and valuable unknown or given information but contain minor errors • Determine the incomplete relevant goal with minor errors • Employ any of the basic concepts with minor error calculation • Represent the problem incompletely with minor errors • Make a prediction (hypothesis), develop criteria, or check existing theory connected to the situation incompletely with minor errors • Formulate incomplete of the reasonable argument, explanation, and solution (including strategy, steps design, and model building) with minor errors • Arrange, choose, evaluate incomplete alternatives with minor errors • Perform the plan incompletely and contain minor errors • Draw incomplete conclusions, evaluation, and reflection on results and methods with minor errors
1 (completely wrong answer and irrelevant OR incomplete answer with major error)	• Identify irrelevant information OR identify incomplete information with major errors • Determine the irrelevant goal OR determine the goal with major errors • Employ the irrelevant basic concept used to solve the problem OR employ any of the basic concept with major errors • Represent the problem irrelevantly OR represent the problem incompletely with major errors • Make a prediction (hypothesis), develop criteria, or check existing theory that are irrelevant and unlogic • Formulate the argument, explanation, and solution that are irrelevant to the problem OR give incomplete answer with major errors • Arrange, choose, and evaluate alternatives that are irrelevant to the problem OR arrange, choose, and evaluate alternatives incompletely with major errors • Perform a plan incorrectly, irrelevant to the problem, and inappropriately • Draw illogic, illegible, or irrelevant conclusions, evaluation, and reflection on results and methods
0	Missing or blank answer

AN INTEGRATED STEM-BASED MATHEMATICAL PROBLEM-SOLVING TEST
(Original Indonesian Version)

GRADE 7

Tes Matematika berbasis STEM (Kelas 7)

Petujuk!

- Kerjakan soal tanpa ada yang dilewatkan
- Terdapat 4 topik dan tiap topik terdiri dari 15 soal
- Pertanyaan tiap topik akan berhubungan satu sama lain (missal nomor 3 di topik "Kemasan Ramah Lingkungan" berkaitan dengan nomor 2, sehingga jika Anda menjawab salah di nomor 2 maka kemungkinan akan salah di nomor 3)
- Sertakan hitungan/cara Anda dalam setiap soal
- Diperbolehkan menggunakan alat bantu, seperti kalkulator
- Waktu pengerjaan tiap topik maksimum 1 jam

Nama :
Kelas :
NISN :
Jenis Kelamin :
Umur :
Sekolah :
Suku :

TOPIK 1: KEMASAN RAMAH LINGKUNGAN

Skenario:

Produk ramah lingkungan dan terjangkau mulai diutamakan. Kami ingin memproduksi sebuah kemasan ramah lingkungan dengan kriteria:

1. Kemasan tersebut memiliki panjang, lebar, dan tinggi 18cm x 18cm x 8cm secara berturut-turut
2. Penutup kemasannya adalah persegi dengan 1,2 kali dari luas alas.
3. Terdapat empat pilihan bahan untuk membuat kemasan:

	Kertas Duplex	**Kertas Ivory**	**Styrofoam**	**Plastik**
Ukuran	79cm×109cm	215mm× 330mm	100cm × 50cm	17cm×50cm
Ketebalan	250gsm	250gsm	0,5cm	0,5cm
Harga	Rp 4.000,00	Rp 2.000,00	Rp 9.000,00	Rp 1.300,00
Waktu yang diperlukan untuk terurai	2-6 bulan		Tidak dapat terurai	1000 tahun

Tantangan:

Desain sebuah kemasan ramah lingkungan dan tentukan harga termurah untuk membuat kemasan tersebut berdasarkan kriteria yang diminta

Pertanyaan:

1. Apa tantangan yang diminta dari skenario tersebut?
2. Apa informasi yang diketahui dan tidak diketahui dalam soal yang sekiranya kamu butuhkan untuk menjawab tantangan tersebut?
3. Berapakah total luas dari keempat sisi kemasan, tutup kemasan, dan alas kemasan dalam cm^2?
4. Apa tebakanmu terkait tipe bahan yang cocok digunakan untuk mendesain kemasan? Beri alasanmu!
5. Gambar desain dari kemasan tersebut berdasarkan bentuk dan ukuran!
6. Sebutkan dua bahan yang tidak mungkin digunakan untuk membuat kemasan tersebut dan beri alasan!
7. Berapa harga pembuatan kemasan berdasarkan bahan yang kamu pilih dan ukuran kemasan?

8. Apakah kemasan yang kamu buat ramah lingkungan dan paling murah? Beri kesimpulan terkait harga dan tipe bahan yang digunakan untuk memproduksi satu kemasan ramah lingkungan!

TOPIK 2: TAMAN SEKOLAH

Skenario:

Sebuah program dari kementerian pendidikan untuk melindungi bumi adalah dengan menanam pohon. Sekolah "Tunas Bangsa" memiliki taman untuk ditanami dengan luas 20m², perbandingan antara lebar dan panjangnya adalah 1:5. Siswa diwajibkan untuk menanam satu jenis tumbuhan disekeliling taman pada Desember 2020. Ada beberapa pilihan tumbuhan yang dapat ditanam dengan jarak ideal disetiap tumbuhan adalah 0,5m. Keefektifan tumbuhan untuk menyerap CO2 akan meningkat $\frac{1}{2}$ kali lipat dari tahun sebelumnya. Berikut beberapa pilihan tumbuhan yang disediakan

	Jumlah CO2 yang diserap (kg/tanaman/tahun)	Diameter (cm)
Annona muricata	75,29	30
Mimusops elengi	34,29	20
Khaya anthotheca	21,29	120

Tantangan:

Desainlah sebuah taman sekolah di "Tunas Bangsa" untuk ditanami satu jenis tumbuhan yang paling efektif untuk menyerap CO2 dan gambarlah grafik yang menunjukkan jumlah total CO2 yang dapat diserap oleh pohon-pohon yang kamu tanam di tahun 2021, 2022, dan 2023

Pertanyaan:

1. Apa tantangan yang diminta dari skenario tersebut?
2. Apa informasi yang diketahui dan tidak diketahui dalam soal yang sekiranya kamu butuhkan untuk menjawab tantangan tersebut?
3. Pohon-pohon tersebut dapat ditanam disekeliling taman. Berdasarkan informasi tersebut, pilihlah konsep luas atau keliling yang dapat dipakai. Berapa luas/keliling taman tersebut dalam m atau m^2 (berdasarkan pilihan konsepmu)?
4. Apa tebakanmu terkait jenis tumbuhan yang akan kamu pilih untuk ditanam? Beri alasanmu!
5. Gambarlah desain dari taman tersebut berdasarkan ukuran, bentuk, posisi pohon, jumlah pohon, diameter pohon, dan jarak antar pohon!
6. Apa satu jenis tumbuhan yang tidak mungkin untuk dipilih? Beri alasanmu!
7. Gunakan kertas berpetak untuk menggambar grafik yang menunjukkan jumlah total dari CO2 yang dapat diserap oleh pohon-pohon yang kamu tanam di tahun 2021, 2022, dan 2023. Gunakan skala yang konsisten untuk menggambar grafik.
8. Apakah jenis tanaman pilihanmu adalah jenis yang paling efektif untuk ditanam berdasarkan kriteria yang diberikan? Berapa jumlah total CO2 yang dapat diserap pohon-pohon ynag ditanam di taman sekolah (dalam kg) pada tahun 2021, 2022, dan 2023 secara berturut-turut?

TOPIK 3: KALORI DAN EMISI GAS RUMAH KACA

Skenario:

Bahan makanan mengeluarkan CO2 selama proses produksi yang terkadang tidak sesuai dengan jumlah kalori yang terkandung dalam bahan makanan tersebut. Kalori rata-rata yang diperlukan pria pada menu utama mereka adalah 1725kkal dan untuk wanita rata-rata sebesar 1500kkal perhari. Tabel dibawah menyajikan informasi tentang kandungan nutrisi dan CO2 yang dihasilkan selama produksi makanan tersebut

Makanan	Kalori (dalam kkal)	Protein (dalam kkal)	Lemak (dalam kkal)	Karbohidrat (dalam kkal)	CO2 yang dihasilkan saat produksi makanan (dalam kg)
Salad	5	2,52	7,2	51,6	1,03
Pepes tahu + nasi	638	56,4	39,6	362,48	2,43
Burger daging keju+ kentang goreng	649	63,2	169,2	360	25,82
Nasi goreng telur ceplok	455	104	249,3	290,4	2,29
Rendang + nasi	841	230	256,5	421,2	25,76
Sate kambing + lontong	548	86,8	128,7	162,4	4,46

Tantangan:

Desainlah "menu harian" mu berdasarkan pilihan menu yang tersedia. Syaratnya adalah:

1. Pilih tiga makanan yang berbeda yang dapat melebihi rata-rata kalori harian yang kamu butuhkan
2. Kandungan total lemak = melebihi 25% dari rata-rata kalori yang dibutuhkan, karbohidrat= melebihi 60% dari rata-rata kalori yang dibutuhkan, dan protein= melebihi 15% dari rata-rata kalori yang dibutuhkan.
3. Setelah memenuhi syarat nomor 1 dan 2, pilihlah makanan dengan pengeluaran CO2 paling sedikit.

Pertanyaan:

1. Apa tantangan yang diminta dari skenario tersebut?
2. Apa informasi yang diketahui dan tidak diketahui dalam soal yang sekiranya kamu butuhkan untuk menjawab tantangan tersebut?
3. Berapa minimal (paling sedikit) kkal lemak, karbohidrat, dan protein yang dibutuhkan untk memenuhi kriteria tersebut?
4. Tebaklah tiga makanan yang dapat kamu pilih sebagai menu harianmu!
5. Beri alasan terkait pilihanmu di nomor 4.
6. Sebutkan dua makanan yang tidak cocok dipilih. Beri alasan!
7. Berapa jumlah total dari masing-masing kalori, protein, lemak, dan karbohidrat dalam tiga makanan pilihanmu (dalam kkal)?
8. Beri kesimpulan dengan menyebutkan tiga makanan yang kamu pilih. Apakah makanan-makanan tersebut memenuhi kriteria dari total kalori, lemak, protein, dan karbohidrat?

Isilah pertanyaan-pertanyaan berikut.

Apa jenis kelamin kamu?	Laki-laki Perempuan
Berada di kelas berapa kamu sekarang?	Kelas 7 Kelas 8 Kelas 9
Umur berapa kamu?	11 tahun 12 tahun 13 tahun 14 tahun 15 tahun 16 tahun
Apa jenis sekolah kamu?	Sekolah negeri Sekolah swasta Sekolah internasional
Di kota/kabupaten manakah letak sekolah kamu?	Surabaya Sidoarjo Lainnya……
Apa suku kamu?	Javanese Madurese Sundanese Other…..
Apa pendidikan terakhir ayah kamu?	SD SMP SMA S1 S2 S3
Apa pendidikan terakhir ibu kamu?	SD SMP SMA S1 S2 S3

Apa pekerjaan ayah kamu?	Dokter Polisi Guru Dosen Insinyur Lainnya….
Apa pekerjaan ibu kamu?	Dokter Polisi Guru Dosen Insinyur Lainnya….
Berapa perkiraan gaji ayah kamu?	Kurang dari IDR 1 juta IDR 1 juta sampai IDR 2,999,000 IDR 3 juta sampai IDR 5 juta Lebih dari 5 juta
Berapa perkiraan gaji ibu kamu?	Kurang dari IDR 1 juta IDR 1 juta sampai IDR 2,999,000 IDR 3 juta sampai IDR 5 juta Lebih dari 5 juta
Ada berapa saudara kandung kamu?	………..

GRADE 8

Tes Matematika berbasis STEM (Kelas 8)

Petujuk!

- Kerjakan soal tanpa ada yang dilewatkan
- Terdapat 4 topik dan tiap topik terdiri dari 15 soal
- Pertanyaan tiap topik akan berhubungan satu sama lain (missal nomor 3 di topik “Penampung Banjir” berkaitan dengan nomor 2, sehingga jika Anda menjawab salah di nomor 2 maka kemungkinan akan salah di nomor 3)
- Sertakan hitungan/cara Anda dalam setiap soal
- Diperbolehkan menggunakan alat bantu, seperti kalkulator
- Waktu pengerjaan tiap topik maksimum 1 jam

Nama	:
Kelas	:
NISN	:
Jenis Kelamin	:
Umur	:
Sekolah	:
Suku	:

TOPIK 1: PENAMPUNG BANJIR

Skenario:

Pemerintah akan membangun bak penampung banjir berbentuk balok dan pompa penyedot di dalamnya di jalan bawah tanah Mayjen. Berikut informasi yang diberikan

1. Volume bak penampung banjir ditentukan dari rata-rata volume air hujan. Rata-rata volume air hujan ditentukan dari rata-rata curah hujan di bulan penghujan tertinggi. Data berikut menunjukkan intensitas curah hujan di bulan penghujan tertinggi, yaitu November (dalam satuan mm/jam) 17, 22, 15, 17, 15, 22, 21, 11, 9, 17, 11, 8, 8, 9, 10, 17, 21, 11, 9, 10
   Catatan: dalam kasus ini 1mm/jam sama dengan 50 liter
2. Ada beberapa pilihan paket pompa untuk menyedot total air hujan yang tertampung di bak penampung. Pilihlah satu paket yang paling efektif untuk menyerap air hujan secara cepat

	Banyak pompa	Jumlah air yang dapat disedot tiap jam oleh satu pompa
Paket 1	4	30liter
Paket 2	3	25liter
Paket 3	2	70liter

Tantangan:

Desain sebuah bak penampung air hujan (dan pompa di dalamnya) dan tentukan waktu tercepat untuk sebuah pilihan paket pompamu menyerap seluruh volume air yang tertampung di bak penampung.

Pertanyaan:

1. Apa tantangan yang diminta dari skenario tersebut?
2. Apa informasi yang diketahui dan tidak diketahui dalam soal yang sekiranya kamu butuhkan untuk menjawab tantangan tersebut?
3. Berapa rata-rata volume air hujan (dalam liter) di bulan November? Gunakan table/grafik (pilih yang paling cocok) untuk menyajikan data intensitas air hujan
4. Tebaklah sebuah paket pompa yang paling cocok untuk diletakkan di bak penampung banjir tersebut. Beri alasanmu!
5. Desainlah bak penampung air hujan (ukuran & bentuk) beserta pompa didalamnya!
6. Sebutkan paket yang tidak mungkin untuk diletakkan di bak penampung air hujan. Beri alasanmu!
7. Berapa waktu yang diperlukan (dalam jam) sebuah paket pilihanmu untuk menyerap total volume di bak penampung tanpa ada yang tersisa?

8. Apakah paket pompa pilihanmu adalah yang paling efektif untuk menyerap air? Beri kesimpulanmu, berapa jam yang diperlukan paket tersebut untuk menyerap total volume air di bak penampung?

TOPIK 2: TAMAN KOTA

Skenario:

SIER adalah salah satu daerah dengan tingkat polusi tinggi di Surabaya. Rata-rata pengeluaran CO2 dari transportasi diprediksi berjumlah 4 juta kg, 4.2 juta kg, dan 4.4 juta kg di tahun 2021, 2022, dan 2023 secara berturut-turut. Kita akan menanam sebuah jenis pohon di taman berbentuk persegi panjang di SIER dengan luas 21m² pada Desember 2020. Perbandingan sisi-sisnya adalah 7:3. Ada beberapa pilihan jenis tananman yang dapat ditanam dengan jarak ideal tiap pohon adalah 50cm.

	Banyak CO2 yang diserap (kg/pohon/tahun)	Diameter
Samanea saman	28.448,39	15 meter
Cassia sp	5.295,47	1 meter
Canangium podratum	756,59	1 meter

Tantangan:

Desain sebuah taman berbentuk persegi panjang untuk ditanami sebuah jenis pohon yang paling efektif menyerap CO2 dan gambarlah grafik dari sisa pengeluaran CO2 oleh transportasi pada tahun 2021, 2022, dan 2023 setelah menanam pohon.

Pertanyaan:

1. Apa tantangan yang diminta dari skenario tersebut?
2. Apa informasi yang diketahui dan tidak diketahui dalam soal yang sekiranya kamu butuhkan untuk menjawab tantangan tersebut?
3. Berapa meter panjang dan lebar dari taman tersebut?
4. Tebaklah satu jenis tanaman yang paling cocok untuk ditanam di taman tersebut. Beri alasanmu!
5. Desainlah taman tersebut beserta ukuran, bentuk, posisi pohon yang akan ditanam, jumlah pohon, dan jarak antar pohon!
6. Sebutkan dua jenis pohon yang tidak mungkin ditanam di taman tersebut. Beri alasanmu!
7. Gunakan kertas berpetak untuk menggambar grafik sisa pengeluaran CO2 oleh transportasi pada tahun 2021, 2022, dan 2023 setelah menanam pohon! Gunakan skala yang konsisten untuk menggambar grafik tersebut.
8. Apakah jenis tanaman yang kamu pilih adalah yang paling efektif untuk ditanam di taman tersebut? Beri kesimpulan terkait berapa kg sisa pengeluaran CO2 oleh transportasi pada tahun 2021, 2022, dan 2023 setelah menanam pohon.

TOPIK 3: KALORI DAN EMISI GAS RUMAH KACA

Skenario:

Bahan makanan mengeluarkan CO2 selama proses produksi yang terkadang tidak sesuai dengan jumlah kalori yang terkandung dalam bahan makanan tersebut. Kalori rata-rata yang diperlukan pria pada menu utama mereka adalah 1725kkal dan untuk wanita rata-rata sebesar 1500kkal perhari. Tabel dibawah menyajikan informasi tentang kandungan nutrisi dan CO2 yang dihasilkan selama produksi makanan tersebut

Makanan	Kalori (dalam kkal)	Protein (dalam kkal)	Lemak (dalam kkal)	Karbohidrat (dalam kkal)	CO2 yang dihasilkan saat produksi makanan (dalam kg)
Salad	5	2,52	7,2	51,6	1,03
Pepes tahu + nasi	638	56,4	39,6	362,48	2,43
Burger daging keju+ kentang goreng	649	63,2	169,2	360	25,82
Nasi goreng telur ceplok	455	104	249,3	290,4	2,29
Rendang + nasi	841	230	256,5	421,2	25,76
Sate kambing + lontong	548	86,8	128,7	162,4	4,46

Tantangan:

Desainlah "menu harian" mu berdasarkan pilihan menu yang tersedia. Syaratnya adalah:

1. Pilih tiga makanan yang berbeda yang dapat melebihi rata-rata kalori harian yang kamu butuhkan
2. Kandungan total lemak = melebihi 25% dari rata-rata kalori yang dibutuhkan, karbohidrat= melebihi 60% dari rata-rata kalori yang dibutuhkan, dan protein= melebihi 15% dari rata-rata kalori yang dibutuhkan.
3. Setelah memenuhi syarat nomor 1 dan 2, pilihlah makanan dengan pengeluaran CO2 paling sedikit.

Pertanyaan:

1. Apa tantangan yang diminta dari skenario tersebut?
2. Apa informasi yang diketahui dan tidak diketahui dalam soal yang sekiranya kamu butuhkan untuk menjawab tantangan tersebut?
3. Berapa minimal (paling sedikit) kkal lemak, karbohidrat, dan protein yang dibutuhkan untk memenuhi kriteria tersebut?
4. Tebaklah tiga makanan yang dapat kamu pilih sebagai menu harianmu!
5. Beri alasan terkait pilihanmu di nomor 4.
6. Sebutkan dua makanan yang tidak cocok dipilih. Beri alasan!
7. Berapa jumlah total dari masing-masing kalori, protein, lemak, dan karbohidrat dalam tiga makanan pilihanmu (dalam kkal)?
8. Beri kesimpulan dengan menyebutkan tiga makanan yang kamu pilih. Apakah makanan-makanan tersebut memenuhi kriteria dari total kalori, lemak, protein, dan karbohidrat?

GRADE 9

Tes Matematika berbasis STEM (Kelas 9)

Petujuk!

- Kerjakan soal tanpa ada yang dilewatkan
- Terdapat 3 topik dan tiap topik terdiri dari 8 soal
- Pertanyaan tiap topik akan berhubungan satu sama lain (misal nomor 3 di topik "Sumur serapan" berkaitan dengan nomor 2, sehingga jika Anda menjawab salah di nomor 2 maka kemungkinan akan salah di nomor 3)
- Sertakan hitungan/cara Anda dalam setiap soal
- Diperbolehkan menggunakan alat bantu, seperti kalkulator
- Waktu pengerjaan tiap topik maksimum 1 jam

Nama	:
Kelas	:
NISN	:
Jenis Kelamin	:
Umur	:
Sekolah	:
Suku	:

TOPIK 1: SUMUR RESAPAN

Skenario:

Salah satu usaha untuk meningkatkan ketersediaan air bersih adalah membangun sumur resapan. Kamu memiliki tugas untuk membangun sumur resapan di rumah Pak Adi. Berikut aturan pembangunan sumur resapan:

1. Sumur tersebut memiliki jarak 1m dari talang rumah.
2. Kamu perlu menambahkan 0.4m untuk membelokkan pipa ke bawah tanah (pada gambar diketahui bahwa akhir dari talang pipa adalah di atas permukaan tanah)
3. Sumur memiliki diameter 100cm, kedalaman 1.5m
4. Untuk menguatkan dinding sumur, digunakan bata (ukuran bata pada satu sisi = 10x22 cm, tiap 40 bata dibutuhkan 11.5kg pasir, perbandingan antar pasir dan semen adalah 4:1)
5. Terdapat saluran masuk air yang mengalirkan air hujan dari pipa talang rumah ke sumur resapan menggunakan pipa paralon
6. Terdapat saluran keluar air dari sumur resapan ke got pembuangan. Saluran keluar air berfungsi membuang kelebihan air pada sumur resapan. Pipa saluran keluar air harus lebih tinggi daripada air got.
7. Harga bahan-bahan yang dibutuhkan untuk membangun sumur resapan adalah sebagai berikut

Bahan-bahan	Harga
Bata	Rp 725/biji
Pipa paralon	Rp 24.000/meter
Semen	Rp 61.000/pcs (40kg)
Pasir	Rp 3200/10 kg

Tantangan:

Desainlah sumur resapan tersebut dan tentukan total harga bahan-bahan untuk membuat sumur resapan tersebut.

Pertanyaan:

1. Apa tantangan yang diminta dari skenario tersebut?
2. Apa informasi yang diketahui dan tidak diketahui dalam soal yang sekiranya kamu butuhkan untuk menjawab tantangan tersebut?
3. Berapa banyak bata, pipa (dalam meter), semen (dalam kg), dan pasir (dalam kg) yang dibutuhkan untuk membangun sumur Pak Adi? Catatan: gunakan 10cm sebagai tinggi bata dan 22cm sebagai panjang bata.
4. Tebaklah dimana posisi dari sumur resapan tersebut?
5. Gambarlah desain sumur resapan tersebut berdasarkan jarak, bentuk, dan ukuran yang diberikan!
6. Berapa meter posisi sumur resapan dari got dan talang?

7. Berapa total harga bahan-bahan untuk membangun sumur resapan berdasarkan desain yang kamu buat?
8. Apakah kamu membeli lebih banyak pipa, semen, dan pasir dari yang kamu butuhkan? Beri kesimpulan terkait total harga bahan-bahan untuk membangun sumur resapan!

TOPIK 2: TAMAN KOTA

Skenario:

SIER adalah salah satu daerah dengan tingkat polusi tinggi di Surabaya. Rata-rata pengeluaran CO2 dari transportasi diprediksi berjumlah 4 juta kg, 4.2 juta kg, dan 4.4 juta kg di tahun 2021, 2022, dan 2023 secara berturut-turut. Kita akan menanam sebuah jenis pohon di taman berbentuk persegi panjang di SIER dengan luas 21m^2 pada Desember 2020. Perbandingan sisi-sisinya adalah 7:3. Ada beberapa pilihan jenis tananman yang dapat ditanam dengan jarak ideal tiap pohon adalah 50cm.

	Banyak CO2 yang diserap (kg/pohon/tahun)	Diameter
Samanea saman	28.448,39	15 meter
Cassia sp	5.295,47	1 meter
Canangium podratum	756,59	1 meter

Tantangan:

Desain sebuah taman berbentuk persegi panjang untuk ditanami sebuah jenis pohon yang paling efektif menyerap CO2 dan gambarlah grafik dari sisa pengeluaran CO2 oleh transportasi pada tahun 2021, 2022, dan 2023 setelah menanam pohon.

Pertanyaan:

1. Apa tantangan yang diminta dari skenario tersebut?
2. Apa informasi yang diketahui dan tidak diketahui dalam soal yang sekiranya kamu butuhkan untuk menjawab tantangan tersebut?
3. Berapa meter panjang dan lebar dari taman tersebut?
4. Tebaklah satu jenis tanaman yang paling cocok untuk ditanam di taman tersebut. Beri alasanmu!
5. Desainlah taman tersebut beserta ukuran, bentuk, posisi pohon yang akan ditanam, jumlah pohon, dan jarak antar pohon!
6. Sebutkan dua jenis pohon yang tidak mungkin ditanam di taman tersebut. Beri alasanmu!
7. Gunakan kertas berpetak untuk menggambar grafik sisa pengeluaran CO2 oleh transportasi pada tahun 2021, 2022, dan 2023 setelah menanam pohon! Gunakan skala yang konsisten untuk menggambar grafik tersebut.
8. Apakah jenis tanaman yang kamu pilih adalah yang paling efektif untuk ditanam di taman tersebut? Beri kesimpulan terkait berapa kg sisa pengeluaran CO2 oleh transportasi pada tahun 2021, 2022, dan 2023 setelah menanam pohon.

TOPIK 3: KALORI DAN EMISI GAS RUMAH KACA

Skenario:

Bahan makanan mengeluarkan CO2 selama proses produksi yang terkadang tidak sesuai dengan jumlah kalori yang terkandung dalam bahan makanan tersebut. Kalori rata-rata yang diperlukan pria pada menu utama mereka adalah 1725kkal dan untuk wanita rata-rata sebesar 1500kkal perhari. Tabel dibawah menyajikan informasi tentang kandungan nutrisi dan CO2 yang dihasilkan selama produksi makanan tersebut

Makanan	Kalori (dalam kkal)	Protein (dalam kkal)	Lemak (dalam kkal)	Karbohidrat (dalam kkal)	CO2 yang dihasilkan saat produksi makanan (dalam kg)
Salad	5	2,52	7,2	51,6	1,03
Pepes tahu + nasi	638	56,4	39,6	362,48	2,43
Burger daging keju+ kentang goreng	649	63,2	169,2	360	25,82
Nasi goreng telur ceplok	455	104	249,3	290,4	2,29
Rendang + nasi	841	230	256,5	421,2	25,76
Sate kambing + lontong	548	86,8	128,7	162,4	4,46

Tantangan:

Desainlah "menu harian" mu berdasarkan pilihan menu yang tersedia. Syaratnya adalah:

1. Pilih tiga makanan yang berbeda yang dapat melebihi rata-rata kalori harian yang kamu butuhkan
2. Kandungan total lemak = melebihi 25% dari rata-rata kalori yang dibutuhkan, karbohidrat= melebihi 60% dari rata-rata kalori yang dibutuhkan, dan protein= melebihi 15% dari rata-rata kalori yang dibutuhkan.
3. Setelah memenuhi syarat nomor 1 dan 2, pilihlah makanan dengan pengeluaran CO2 paling sedikit.

Pertanyaan:

1. Apa tantangan yang diminta dari skenario tersebut?
2. Apa informasi yang diketahui dan tidak diketahui dalam soal yang sekiranya kamu butuhkan untuk menjawab tantangan tersebut?
3. Berapa minimal (paling sedikit) kkal lemak, karbohidrat, dan protein yang dibutuhkan untk memenuhi kriteria tersebut?
4. Tebaklah tiga makanan yang dapat kamu pilih sebagai menu harianmu!
5. Beri alasan terkait pilihanmu di nomor 4.
6. Sebutkan dua makanan yang tidak cocok dipilih. Beri alasan!
7. Berapa jumlah total dari masing-masing kalori, protein, lemak, dan karbohidrat dalam tiga makanan pilihanmu (dalam kkal)?
8. Beri kesimpulan dengan menyebutkan tiga makanan yang kamu pilih. Apakah makanan-makanan tersebut memenuhi kriteria dari total kalori, lemak, protein, dan karbohidrat?

APPENDIX B

MATHEMATICS DSPK TEST
(English Version)

The Construct of the Test

Format	Topics or domain of knowledge	Total of items	Items
Multiple-choice	Integer and fraction (Meaning, conversion, ordering)	9	1—9
	Measurement (approximation)	4	10—13
	Ratio & proportion (velocity, inverse and direct proportion)	8	14—21
	Geometry (length, 2D figure, circle)	4	22—25
	Statistics (mean)	5	26—30
	Knowledge of meaning	7	1, 3, 10, 14, 18, 21, 26
	Integration of knowledge	10	2, 5, 7, 9, 11, 19, 23, 25, 27, 30
	Application of knowledge	13	4, 6, 8, 12, 13, 15-17, 20, 22, 24, 28, 29

The Instrument

Planting Activity

The Surabaya government has a greenery area. They want to plant *Samanea saman*, *Cassia sp*, and *Canangium podratum* in the area without remaining. The criteria are:

1. Cassia sp has bigger partition than Samanea saman
2. Canangium podratum has smaller partition than Samanea saman
3. The partition of Samanea saman is $\frac{1}{4}$.

1 What is the possible arrangement from the smallest partition to the biggest partition based on the information above?
   - A. *Cassia Sp*, *Samanea saman*, and *Canangium podratum*
   - **B. *Canangium podratum*, *Samanea saman*, and *Cassia sp***
   - C. *Samanea saman, Canangium podratum, and Cassia sp*
   - D. *Samanea saman, Cassia sp, and Canangium podratum*

2 How does your design if you're asked to make a partition based on the information above?
   - **A. An area is divided into three partitions. Canangium podratum has the smallest partition, and Cassia sp has the biggest partition**
   - B. An area is divided into three partitions. Cassia sp has the smallest partition, and Canangium podratum has the biggest partition
   - C. An area is divided into three partitions. Samanea saman has the smallest partition, and Cassia sp has the biggest partition
   - D. There are more than one area and divided into three partitions

3 What does the meaning of $\frac{1}{4}$ that represent the partition of Samanea saman?
   - **A. There are four parts of a 1, and $\frac{1}{4}$ equals one of those 4 parts***
   - B. There are five parts of a 1, and $\frac{1}{4}$ equals one of those 5 parts
   - C. There is a part, and $\frac{1}{4}$ equals four of one part
   - D. There is a part from a whole five parts

4 What are the decimal and percentage that represent the partition of Samanea saman respectively?
   - A. 0.025; 25%
   - B. 0.45; 45%
   - **C. 0.25; 25%**
   - D. 0.25; 250%

5 What is the possible ordering representing the area partition of *Canangium podratum*, *Samanea saman*, and *Cassia sp*, respectively?
   - A. $0.2;\ 0.45;\ 0.6$
   - **B. $\mathbf{0.125;\ 0.25;\ 0.625}$***
   - C. $0.25;\ 0.125;\ 0.625$
   - D. $0.625;\ 0.25;\ 0.125$

CO2 Emission

The CO2 emission from transportation in Surabaya on April 2021 is 2396 kg/hour. It becomes 488 kg/hour on Mei but only 75% is from transportation. The CO2 emission in June decreases 2.45 kg/hour from the total transportation CO2 emission in April and Mei.

6 What is the amount of CO2 emission from transportation on Mei?
   - A. 3660kg/hour
   - B. 36600kg/hour
   - **C. 366kg/hour***
   - D. 488kg/hour

7 How do you calculate the amount of CO_2 emission from transportation in June?

A. 2396+(488×75%)−2.45*

B. (2396+488) × (75%−2.45)

C. (2396+488)×75%−2.45

D. 2396+488−2.45

8 How many CO_2 emissions from transportation in June?

A. 2759.55*

B. 2160.55

C. 5.212

D. -4758.6

9 Choose the correct statement related to the amount of CO_2 emission from transportation in June.

A. The amount of CO_2 emission from transportation in June is smaller than in April, with 7154.6 kg/hour in difference

B. The amount of CO_2 emission from transportation in June is bigger than in April, with 2816 kg/hour in difference

C. The amount of CO_2 emission from transportation in June is smaller than in April, with 235.45 kg/hour in difference

D. The amount of CO_2 emission from transportation in June is bigger than in April, with 363.55 kg/hour in difference*

Car Parking

You want to design an outdoor car parking area in a $30m \times 15m$. The size of the biggest car in Indonesia is $4915mm \times 2155mm$, and the smallest car is $3430\ mm \times 1460mm$. You are asked to calculate the maximum number of possible cars to fulfil the area with all types of cars could be fitted and without disturbing the mobility of cars

10 What kind of car size do you need to consider if you want to calculate the maximum number of possible cars to fulfill the area and consider how all types of car sizes can be fitted?

A. Biggest car size

B. Smallest car size

C. Both

D. None of them

11 Based on information from the 'car parking' question, what kind of parking area design can you create to fulfill the maximum number of cars without disturbing car mobility, and all types of car sizes can be fitted?

A. Rectangle with the entrance in the middle of 30m side. There are two rows of cars, one of them is on the left, and the other is on the right sides of the entrance. The car's position is parking uprightly (up-down)

B. Rectangle with the entrance in the middle of 15m side. There are two rows of cars, one of them is on the left, and the other is on the right sides of the entrance. The car's position is parking uprightly (up-down)

C. Rectangle with the entrance in the middle of 30m side. There are two rows of cars, one of them is on the left, and the other is on the right sides of the entrance. The car's position is parking parallelly (right-left)

D. Rectangle with the entrance in the middle of 15m side. There are two rows of cars, one of them is on the left, and the other is on the right sides of the entrance. The car's position is parking parallelly (right-left)

12 What is the approximation area for parking a car based on your design?

A. $4.915m \times 2.155m$

B. $3.430\ m \times 1.460m$

C. $4m \times 2m$

D. $5m \times 3m$

13 Based on the design that you choose; how many maximum cars are possible to fulfill this area?
A. 90 cars
B. 20 cars*
C. 49 cars
D. 70 cars

Covid-19 Vaccination Car

A car for dropping the Covid-19 vaccine in Surabaya has to arrive in a hospital at 6.30 a.m West Indonesian Time. A driver plans to ride his car at an average speed of 50km/hour. The distance between the Covid vaccination center and the hospital is 7500m. He needs to consider the unexpected time spending for waiting several traffic lights (there are 6 traffic lights with 60 seconds in every traffic light).

14 What information do you need to determine the departure time for driver to go to the hospital?
A. Average speed, the distance between Covid vaccination center and the hospital, and arrival time
B. Average speed, the distance between Covid vaccination center and the hospital
C. Average speed, the distance between Covid vaccination center and the hospital, time for several traffic light, and arrival time
D. Average speed, the distance between Covid vaccination center and the hospital, and time for several traffic light

15 Convert the distance between the Covid vaccination center and the hospital into kilometer
A. 750
B. 0.75
C. 7.5*
D. 75

16 How many hours that he needs to drive the car from the Covid vaccination center to the hospital?
A. 0.15
B. 0.25
C. 2.5
D. 0.4

17 What is the best time for him to go to the hospital?
A. Before 6.15 a.m.*
B. 6.21 a.m.
C. Before 4.00 a.m.
D. After 5.59 a.m.

Raising Cows

A cow has some benefits for completing human protein needs. However, there are several dilemmas related to the environment: (1) cows need grass for eating and it can destroy the environment; (2) cows produce methane gas (CH4) from their digestive process. The following table represents the day for cows to consume 1 ton of grass and CH4 production per month (in kg) by cows.

The number of cows	The day spent for cows to consume 1 ton of grass	Methane gas production/month (kg)
25	4	125
50	2	250
100	1	500

18 $\frac{1}{5}$ is represented the ratio of the number of cows and their methane gas production per month (in kg). What is the correct meaning of the statement?

A. For every five cows produce 1kg of methane gas per month

B. For every cow produces 5kg of methane gas per month*

C. There are 5kg of methane gas per month

D. There is only a cow that produces 5kg of methane gas per month

19 Choose the correct statement based on the table if you are asked about how many days that 10 cows needed to consume 1-ton grass and what is the amount of methane gas produced by them per month (in kg)

A. The day needed for 10 cows to consume 1-ton grass will decrease than bigger number of cows. The methane gas produced per month by 10 cows will increase

B. The day needed for 10 cows to consume 1-ton grass will increase than bigger number of cows. The methane gas produced per month by 10 cows will increase

C. The day needed for 10 cows to consume 1-ton grass will decrease than bigger number of cows. The methane gas produced per month by 10 cows will decrease

D. The day needed for 10 cows to consume 1-ton grass will increase than bigger number of cows. The methane gas produced per month by 10 cows will decrease*

20 How many days that 10 cows needed to consume 1 ton of grass and the amount of methane gas produced per month (in kg), respectively?

A. 1.6 and 5000

B. 50 and 10

C. 10 and 50*

D. 12.5 and 5000

21 Choose the correct statement based on the table above

A. Because the multiplication of the number of cows and the day needed to consume 1 ton grass is 100. Hence, 10 cows consume 1 ton of grass without remaining in 10 days*

B. Because the ratio of a cow and its methane gas production per month is 1:5. Hence, the amount of methane gas production by 10 cows is 5 kg per month

C. Because the ratio of the number of cows and their methane gas production per month is 1:100. Hence, the amount of methane gas production by 10 cows is 5000kg per month

D. Because the ratio of the number of cows and the day spent to consume 1 ton grass is 1:10. Hence, 10 cows consume 1 ton of grass without remaining in 10 days

Greenery Areas

A greening activity can be used for reducing air pollution. One of the programs in the city is opening a greenery area. There are two types of greenery areas in Surabaya. These greenery areas could only be planted around them with a tree that has a diameter of 30cm and the distance of each tree is 0.5 m.

Type 1: A greenery area that has a radius of 7m.

Type 2: A rectangle greenery area that has an area of $20m^2$, $length > width$

22 What is the possible length and width of the type 2 area, respectively if $l > w$?

A. 6m and 4m

B. 10m and 10m

C. 10m and 2m*

D. 4m and 5m

23 How is the possible design of the type 1 greenery area based on the information above?

A. **A circle greenery area with the 0.3 diameter trees around it. The distance between each tree is 0.5m**

B. A circle greenery area with the 0.3 diameter trees evenly distributed in it. The distance between each tree is 0.5m.

C. A circle greenery area with the trees evenly distributed in it

D. A square greenery area with the 0.3 diameter trees around it. The distance between each tree is 0.5m.

24 How many maximum trees are possible to be planted in the “type 1” greenery area?

A. 30 trees

B. 55 trees*

C. 192 trees

D. 308 trees

25 Choose the correct statements based on given situations in a “Greenery Areas” word problem.

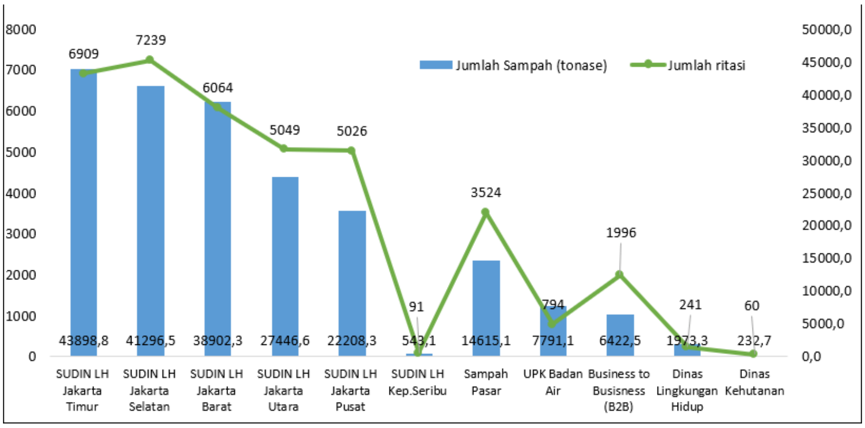
A. The number of trees that could be planted in the “type 1” greenery area is less than the number of trees that could be planted in the “type 2” greenery area

B. The number of trees that could be planted in the “type 1” greenery area is the same as the number of trees that could be planted in the “type 2” greenery area

C. The difference between the number of threes that could be planted in the “type 1” greenery area and the number of trees that could be planted in the “type 2” greenery area is 7

D. **The number of trees that could be planted in the “type 1” greenery area is more than the number of trees that could be planted in the “type 2” greenery area***

Waste in a landfill



-Average amount of waste transported by a truck from the “B2B”= 3.2

-Average amount of waste transported by a truck form the “Dinas Lingkungan Hidup”= 8.2

-Average amount of waste transported by a truck from the “Dinas Kehutanan”=3.8

PS: Jumlah ritasi: the number of waste trucks that are coming to the landfills

Source: statistik.jakarta.go.id

The graph explains the total waste and the total number of trucks that come to Jakarta’s landfill from several waste dumps on October.

26 Look at these statements:

1. The bar chart represents the total amount of waste in every waste dump
2. The line chart represents the number of the truck that is coming to the landfills from waste dumps
3. The bar chart represents the number of the truck that is coming to the landfills from waste dumps
4. The line chart represents the total amount of waste in every waste dump

Which of these following statements are correct based on the graph.

A. 1 and 4

B. 2 and 3

C. 1 and 2*

D. 3 and 4

27 If you are asked to calculate the mean of waste transported by a truck on October to the landfill, what informations do you need to be calculated?

A. The total amount of waste in the landfill and the total number of waste trucks that come to the landfill*

B. The number of waste dumps and the total amount of waste in the landfill

C. The number of waste dumps and the total number of waste trucks that come to the landfill

D. The number of waste dumps, the total number of waste trucks that come to the landfill, and the total amount of waste in the landfill

28 What is the total amount of waste (tons) in the landfill and the total truck that come to the landfill, respectively?

A. 36990 and 205330,3

B. 36993 and 205330,3

C. 205330 and 36993

D. 205330,3 and 36993*

29 How many tons of waste in average is transported by a truck to the landfill on October?

A. 5.5

B. 5.6*

C. 0.18

D. 5.0

30 Choose the correct statement based on the information from the graph related to the average amount of waste transported by a truck to the landfill on October!

A. The difference between the average amount of waste transported by a truck on October to the landfill and the average amount of waste transported by a truck from B2B is 2.3 tons

B. The average amount of waste transported by a truck on October to the landfill is smaller than the average amount of waste transported by a truck from the "Dinas Lingkungan Hidup" waste dump, with 8.02 tons in difference

C. The difference between the average amount of waste transported by a truck on October to the landfill and the average amount of waste transported by a truck from the "Dinas Kehutanan" is 1.2 tons

D. The average amount of waste transported by a truck on October to the landfill is bigger than the average amount of waste transported by a truck from the "B2B" waste dump, with 2.4 tons in difference *

<u>MATHEMATICS DSPK TEST</u>
<u>(Original Indonesian Version)</u>
Tes Pengetahuan Dasar Matematika

Petujuk!

- Kerjakan dengan jujur dan serius tanpa ada yang dilewatkan
- Terdapat 30 soal pilihan ganda (dengan 4 pilihan jawaban)
- Pilih dengan meng "klik" jawaban yang menurut Anda benar
- Total waktu pengerjaan adalah 2 jam pelajaran
- Tidak diperbolehkan menggunakan alat bantu apapun (kalkulator aplikasi online, dll)
- Terdapat soal yang berhubungan satu sama lain (missal soal nomor 2 berhubungan dengan soal nomor 1, jadi jika Anda salah menjawab di nomor 1 kemungkinan juga akan salah menjawab di nomor 2)
- Jika sudah menjawab semua soal, klik "kirim" untuk mengirimkan jawaban Anda

Selamat Mengerjakan!

Nama	:
Kelas	:
NISN	:
Jenis Kelamin	:
Umur	:
Sekolah	:
Suku	:

Aktivitas Penghijauan (Soal Nomor 1-5)

Pemerintah kota Surabaya memiliki sebuah lahan penghijauan. Mereka ingin menanam *Samanea saman*, *Cassia sp*, dan *Canangium podratum* di area tersebut tanpa bersisa. Mereka menetapkan kriteria sebagai berikut:

1. Cassia sp memiliki bagian lebih besar dari Samanea saman
2. Canangium podratum memiliki bagian lebih kecil dari Samanea saman
3. Bagian yang ditetapkan untuk Samanea saman adalah $\frac{1}{4}$.

1 Bagaimanakah urutan yang benar dari bagian terkecil ke bagian terbesar berdasarkan informasi diatas?
   A. *Cassia Sp*, *Samanea saman*, dan *Canangium podratum*
   B. *Canangium podratum*, *Samanea saman*, dan *Cassia sp*
   C. *Samanea saman, Canangium podratum, dan Cassia sp*
   D. *Samanea saman, Cassia sp, dan Canangium podratum*

2 Bagaimana desain yang akan kamu rencanakan jika kamu diminta untuk menentukan bagian yang akan ditanami pohon-pohon tersebut berdasarkan informasi diatas?
   A. Satu lahan dibagi menjadi tiga bagian. Canangium podratum memiliki bagian terkecil dan Cassia sp memiliki bagian terbesar
   B. Satu lahan dibagi menjadi tiga bagian. Cassia sp memiliki bagian terkecil dan Canangium podratum memiliki bagian terbesar
   C. Satu lahan dibagi menjadi tiga bagian. Samanea saman memiliki bagian terkecil dan Cassia sp memiliki bagian terbesar
   D. Ada lebih dari satu lahan yang dibagi menjadi tiga bagian

3 Apa maksud dari $\frac{1}{4}$ yang menyatakan bagian dari Samanea saman?

A. Ada satu bagian sebanyak empat, dan $\frac{1}{4}$ memiliki makna satu bagian dari empat tersebut
B. Ada satu bagian sebanyak 5, dan $\frac{1}{4}$ memiliki makna satu bagian dari lima tersebut
C. Ada satu bagian, dan $\frac{1}{4}$ memiliki makna empat dari satu bagian tersebut
D. Ada satu bagian dari lima keseluruhannya

4 Bagaimana bentuk decimal dan persen yang menyatakan bagian dari Samanea saman secara berturut-turut?

A. 0.025; 25%
B. 0.45; 45%
C. 0.25; 25%
D. 0.25; 250%

5 Bagaimanakah urutan yang memungkinkan yang menyatakan bagian dari *Canangium podratum*, *Samanea saman*, dan *Cassia sp*, secara berturut-turut?

A. $0.2;\ 0.45;\ 0.6$
B. $0.125;\ 0.25;\ 0.625$
C. $0.25;\ 0.125;\ 0.625$
D. $0.625;\ 0.25;\ 0.125$

Pengeluaran CO2 (Soal Nomor 6-9)

Pengeluaran CO2 dari transportasi di Surabaya pada bulan April sebesar 2396 kg/jam. Pada bulan Mei pengeluaran CO2 secara keseluruhan menjadi 488 kg/jam. Dari jumlah tersebut 75% nya disebabkan oleh transportasi. Pengeluaran CO2 di bulan Juni menurun sebesar 2.45 kg/jam dari total pengeluaran CO2 yang disebabkan oleh transportasi di bulan April dan Mei.

6 Berapakah jumlah pengeluaran CO2 dari transportasi pada bulan Mei?

A. 3660kg/jam
B. 36600kg/jam
C. 366kg/jam
D. 488kg/jam

7 Bagaimana kamu menghitung jumlah pengeluaran CO2 dari trasportasi pada bulan Juni?

A. 2396+(488×75%)−2.45
B. (2396+488) × (75%−2.45)
C. (2396+488)×75%−2.45
D. 2396+488−2.45

8 Berapakah pengeluaran CO2 dari trasportasi pada bulan Juni?

A. 2759.55
B. 2160.55
C. 5.212
D. -4758.6

9 Pilih pernyataan yang benar berdasarkan jumlah pengeluaran CO2 yang disebabkan oleh transportasi.

A. Jumlah pengeluaran CO2 yang disebabkan transportasi pada bulan Juni lebih sedikit daripada bulan April, dengan selisih sebesar 7154.6 kg/jam
B. Jumlah pengeluaran CO2 yang disebabkan transportasi pada bulan Juni lebih besar dibanding bulan April, dengan selisih sebesar 2816 kg/jam
C. Jumlah pengeluaran CO2 yang disebabkan transportasi pada bulan Juni lebih sedikit dibanding bulan April, dengan selisih sebesar 235.45 kg/jam
D. Jumlah pengeluaran CO2 yang disebabkan transportasi pada bulan Juni lebih besar dibanding bulan April, dengan selisih sebesar 363.55 kg/jam

Parkiran Mobil (Soal Nomor 10-13

Kamu akan mendesain sebuah parkiran mobil terbuka di sebuah lahan sebesar $30m \times 15m$. Ukuran mobil terbesar di Indonesia adalah $4915mm \times 1855mm$, dan ukuran mobil terkecil adalah $3430\ mm \times 1460mm$. Kamu diminta untuk menghitung berapa jumlah mobil terbanyak yang mungkin untuk memenuhi lahan tersebut.

10 Ukuran mobil manakah yang akan kamu utamakan jika kamu akan menghitung jumlah mobil terbanyak yang dapat memenuhi area parkir tersebut dan memikirkan semua jenis mobil dapat masuk ke area parkir?

A. Ukuran mobil terbesar
B. Ukuran mobil terkecil
C. Keduanya
D. Tidak keduanya

11 Berdasarkan informasi tentang "parkiran mobil", jenis area parkir seperti apa yang akan kamu buat untuk memaksimalkan penggunaan area parkir tersebut dan tanpa mengganggu pergerakan mobil nantinya?

A. Persegi Panjang dengan pintu masuk ditengah sisi 30m. Terdapat dua baris mobil, satu di sisi kanan dan satu di sisi kiri pintu masuk. Posisi mobil tersebut parkir secara vertikal
B. Persegi Panjang dengan pintu masuk ditengah sisi 15m. Terdapat dua baris mobil, satu di sisi kanan dan satu di sisi kiri pintu masuk. Posisi mobil tersebut parkir secara vertikal
C. Persegi Panjang dengan pintu masuk ditengah sisi 30m. Terdapat dua baris mobil, satu di sisi kanan dan satu di sisi kiri pintu masuk. Posisi mobil tersebut parkir secara horisontal
D. Persegi Panjang dengan pintu masuk ditengah sisi 15m. Terdapat dua baris mobil, satu di sisi kanan dan satu di sisi kiri pintu masuk. Posisi mobil tersebut parkir secara horisontal

12 Berapakah taksiran lahan yang dapat digunakan untuk memarkir satu mobil berdasarkan desain yang kamu buat?

A. $4.915m \times 1.855m$
B. $3.430\ m \times 1.460m$
C. $4m \times 2m$
D. $5m \times 3m$

13 Berdasarkan desain yang kamu pilih, berapa jumlah mobil terbanyak yang dapat memenuhi lahan parkir tersebut?

A. 90 mobil
B. 20 mobil
C. 49 mobil
D. 70 mobil

Mobil Vaksin Covid-19 (Soal Nomor 14-17)

Sebuah mobil untuk mengantar vaksin Covid-19 di Surabaya harus tiba di rumah sakit pada pukul 06.30 WIB. Supir mobil tersebut berencana untuk mengendarai mobilnya dengan kecepatan rata-rata 50km/jam. Jarak dari pusat pengadaan vaksin Covid-19 ke rumah sakit adalah 7500m. Dia perlu untuk mempertimbangkan kemacetan yang tak terduga dan waktu untuk menunggu lampu lalu-lintas (ada 6 lampu lalu lintas, tiap lampu lalu lintas membutuhkan waktu 60 detik untuk signal berhenti). Bantu supir tersebut untuk menentukan waktu terbaik dia untuk berangkat dari pusat pengadaan vaksin menuju rumah sakit.

14 Informasi apa yang kamu butuhkan untuk menghitung waktu yang diperlukan supir untuk mengendarai mobilnya menuju rumah sakit?
   A. Kecepatan rata-rata, jarak antara pusat pengadaan vaksin dan rumah sakit, dan waktu tiba di rumah sakit
   B. Kecepatan rata-rata, jarak antara pusat pengadaan vaksin dan rumah sakit
   C. Kecepatan rata-rata, jarak antara pusat pengadaan vaksin dan rumah sakit, waktu untuk menunggu beberapa rambu lalu lintas, dan waktu tiba di rumah sakit
   D. Kecepatan rata-rata, jarak antara pusat pengadaan vaksin dan rumah sakit, dan waktu untuk menunggu beberapa rambu lalu lintas

15 Ubah satuan jarak antara pusat pengadaan vaksin dan rumah sakit dalam satuan kilometer
   A. 750
   B. 0.75
   C. 7.5
   D. 75

16 Berapa jam yang dibutuhkan supir untuk mengendarai mobil dari pusat pengadaan vaksin ke rumah sakit?
   A. 0.15
   B. 0.25
   C. 2.5
   D. 0.4

17 Jam berapakah sebaiknya dia pergi menuju ke rumah sakit?
   A. Sebelum 6.15 WIB
   B. 6.21 WIB
   C. 4.00 WIB
   D. Setelah 5.59 WIB

Beternak Sapi (Soal Nomor 18-21)

Sapi memiliki banyak manfaat terkait pemenuhan kebutuhan protein manusia.Tetapi, ada beberapa hal yang merugikan lingkungan: (1) Sapi membutuhkan rumput untuk makanan utama yang dapat mengganggu keseimbangan ekosistem; (2) Sapi mengeluarkan gas metana (CH4) dari proses pencernaan yang akan mencemari lingkungan. Tabel dibawah ini menjelaskan tentang jumlah hari yang dibutuhkan sapi untuk mengkonsumsi 1 ton rumput dan produksi CH4 per bulan (dalam kg).

Jumlah sapi	Hari yang dibutuhkan untuk menghabiskan 1 ton rumput	Produksi CH4/bulan (kg)
25	4	125
50	2	250
100	1	500

18 $\frac{1}{5}$ menunjukkan rasio dari satu sapi dan jumlah gas metana yang dikeluarkan oleh nya (dalam kg). Manakah maksud yang benar dari pernyataan tersebut?
   A. Untuk setiap 5 sapi, memproduksi 1kg gas metana per bulan
   B. Untuk setiap seekor sapi, memproduksi 5kg gas metana per bulan
   C. Terdapat 5kg gas metana per bulan
   D. Hanya terdapat satu sapi yang memproduksi 5kg gas metana per bulan

19 Pilih pernyataan yang benar berdasarkan table jika kamu diminta untuk mencari berapa hari yang dibutuhkan 10 sapi untuk menghabiskan 1 ton rumput dan berapa jumlah gas metana yang dihasilkan per bulan (dalam kg)

A. Jumlah hari yang dibutuhkan oleh 10 sapi untuk menghabiskan 1 ton rumput akan menurun dibandingkan dengan jumlah hari yang dibutuhkan oleh sapi yang lebih banyak. Gas metana yang dihasilkan oleh 10 sapi akan naik.
B. Jumlah hari yang dibutuhkan oleh 10 sapi untuk menghabiskan 1 ton rumput akan naik dibandingkan dengan jumlah hari yang dibutuhkan oleh sapi yang lebih banyak. Gas metana yang dihasilkan oleh 10 sapi akan naik.
C. Jumlah hari yang dibutuhkan oleh 10 sapi untuk menghabiskan 1 ton rumput akan menurun dibandingkan dengan jumlah hari yang dibutuhkan oleh sapi yang lebih banyak. Gas metana yang dihasilkan oleh 10 sapi akan menurun.
D. Jumlah hari yang dibutuhkan oleh 10 sapi untuk menghabiskan 1 ton rumput akan naik dibandingkan dengan jumlah hari yang dibutuhkan oleh sapi yang lebih banyak. Gas metana yang dihasilkan oleh 10 sapi akan menurun

20 Berapa hari yang dibutuhkan oleh 10 sapi untuk menghabiskan 1 ton rumput dan berapa gas metana yang dihasilkan oleh mereka per bulan (dalam kg), secara berturut-turut?

A. 1.6 dan 5000
B. 50 dan 10
C. 10 dan 50
D. 12.5 dan 5000

21 Pilihlah pernyataan yang benar berdasarkan tabel diatas

A. Karena perkalian dari jumlah sapi dan hari yang dibutuhkan untuk menghabiskan 1 ton rumput adalah 100. Maka, 10 sapi menghabiskan 1 ton rumput dalam 10 hari
B. Karena rasio dari seekor sapi dan gas metana yang dihasilkannya per bulan adalah 1:5. Maka julah gas metana yang dihasilkan oleh 10 sapi adalah 5 kg per bulan.
C. Karena rasio banyak sapi dan gas metana yang dihasilkan per bulan adalah 1:100. Maka, jumlah gas yang dihasilkan 10 sapi adalah 5000 kg per bulan
D. Karena rasio banyak sapi dan hari yang dibutuhkan untuk menghabiskan 1 ton rumput adalah 1:10. Maka, 10 sapi menghabiskan 1 ton rumput dalam 10 hari

Lahan Penghijauan (Soal Nomor 22-25)

Sebuah lahan penghijauan dapat digunakan untuk menurunkan polusi udara. Salah satu programnya yaitu membuka lahan penghijauan. Ada dua jenis lahan penghijauan di Surabaya. Lahan tersebut hanya dapat ditanami pohon disekelilingnya dengan pohon yang memiliki diameter 30cm dan jarak antar pohon 0.5m.

Lahan tipe 1: Lahan penghijauan yang memiliki diameter 7m.

Lahan tipe 2: Lahan penghijauan berbentuk persegi Panjang dengan luas ^{2}, $p > l$

22 Berapakah panjang dan lebar yang mungkin dari lahan tipe 2 tersebut secara berturut-turut jika $p > l$?

A. 6m dan 4m
B. 10m dan 10m
C. 10m dan 2m
D. 4m dan 5m

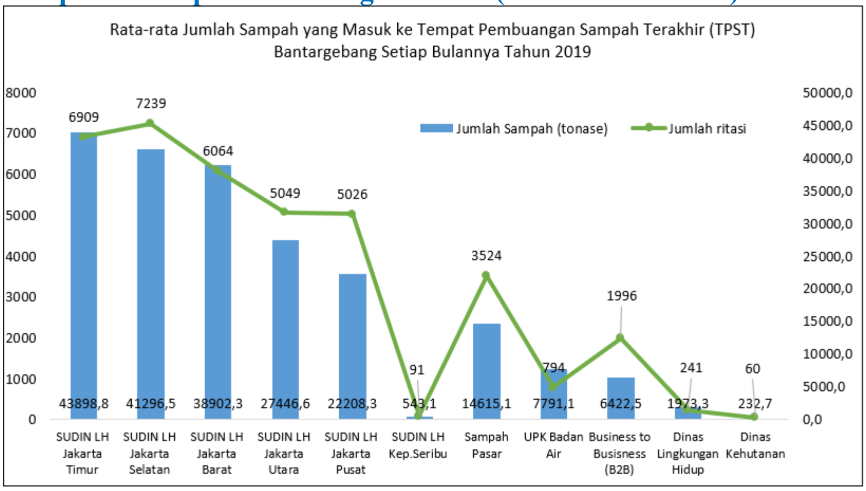
23 Bagaimanakah desain yang memungkinkan dari lahan tipe 1 berdasarkan informasi diatas agar memuat pohon lebih banyak?

A. Lahan lingkaran dengan pohon (diameter 0.5) disekelilingnya
B. Lahan lingkaran dengan pohon (diameter 0.5) yang menyebar diseluruh area
C. Lahan lingkaran dengan pohon yang menyebar diseluruh area
D. Lahan persegi dengan pohon (diameter 0.5) disekelilingnya

24 Berapa jumlah pohon terbanyak yang dapat ditanam di lahan "tipe 1"?
A. 30 pohon
B. 55 pohon
C. 192 pohon
D. 308 pohon

25 Pilihlah pernyataan yang benar berdasarkan informasi diatas
A. Jumlah pohon yang ditanam di lahan "tipe 1" lebih sedikit dari jumlah pohon yang ditanam di lahan "tipe 2"
B. Jumlah pohon yang ditanam di lahan "tipe 1" sama dengan jumlah pohon yang ditanam di lahan "tipe 2"
C. Selisih dari jumlah pohon yang ditanam di lahan "tipe 1" dan jumlah pohon yang ditanam di lahan "tipe 2" adalah 7
D. Jumlah pohon yang ditanam di lahan "tipe 1" lebih dari jumlah pohon yang ditanam di lahan "tipe 2"

Sampah di tempat Pembuangan Akhir (Soal Nomor 26-30)



NB: Jumlah ritasi: jumlah truk sampah yang datang ke tempat pembuangan akhir
Sumber: statistik.jakarta.go.id
Grafik menjelaskan jumlah sampah di Jakarta per bulan di tahun 2019 dari beberapa TPS

26 Perhatikan pernyataan berikut
1. Diagram batang menyatakan total sampah pada setiap tempat pembuangan sementara
2. Diagram garis menyatakan jumlah truk yang datang ke tempat pembuangan akhir dari masing-masing tempat pembuangan sementara
3. Diagram batang menyatakan jumlah truk yang datang ke tempat pembuangan akhir dari masing-masing tempat pembuangan sementara
4. Diagram garis menyatakan total sampah pada setiap tempat pembuangan sementara

Manakah dari pernyataan tersebut yang benar berdasarkan grafik?
A. 1 dan 4
B. 2 dan 3
C. 1 dan 2
D. 3 dan 4

27 Jika kamu diminta untuk menghitung rata-rata jumlah sampah yang diangkut sebuah truk tiap bulan, informasi apa yang kamu butuhkan untuk dihitung?

A. Total seluruh sampah dan total truk yang masuk ke tempat pembuangan akhir
B. Jumlah tempat pembuangan sementara dan total seluruh sampah
C. Jumlah tempat pembuangan sementara dan total truk yang masuk ke tempat pembuangan akhir
D. Jumlah tempat pembuangan sementara, total truk yang masuk ke tempat pembuangan akhir, dan total seluruh sampah

28 Berapakah total sampah yang masuk ke tempat pembuangan akhir (dalam ton) dan berapakah jumlah truk yang masuk ke tempat pembuangan akhir, secara berturut-turut?

A. 36990 dan 205330,3
B. 36993 dan 205330,3
C. 205330 dan 36993
D. 205330,3 dan 36993

29 Berapa ton rata-rata sampah yang diangkut oleh sebuah truk di tempat pembuangan akhir tiap bulan?

A. 5.5
B. 5.6
C. 0.18
D. 5.0

30 Pilihlah pernyataan yang benar berdasarkan grafik yang berhubungan dengan rata-rata jumlah sampah yang diangkut oleh sebuah truk ke tempat pembuangan akhir tiap bulan!

A. Selisih antara rata-rata jumlah sampah yang diangkut oleh sebuah truk ke tempat pembuangan akhir tiap bulan dan rata-rata jumlah sampah yang diangkut dari B2B adalah 2.3 ton
B. Rata-rata jumlah sampah yang diangkut oleh sebuah truk ke tempat pembuangan akhir tiap bulan lebih sedikit dari rata-rata jumlah sampah yang diangkut dari tempat pembuangan sementara “Dinas Lingkungan Hidup” dengan selisih 8.02 ton
C. Selisih antara rata-rata jumlah sampah yang diangkut oleh sebuah truk ke tempat pembuangan akhir tiap bulan dan rata-rata jumlah sampah yang diangkut dari tempat pembuangan sementara “Dinas Kehutanan” adalah 1.2 ton
D. Rata-rata jumlah sampah yang diangkut oleh sebuah truk ke tempat pembuangan akhir tiap bulan lebih banyak dari rata-rata jumlah sampah yang diangkut dari tempat pembuangan sementara “B2B” dengan selisih 2.4 ton

APPENDIX C

TEXT COMPREHENSION TEST
(English Version)

The Construct of the Test

Format	Indicators	Number of items	Items
Multiple-choice, matching and true/false, short answer, long answer	Finding information	6	1, 7, 8, 10, 16, 20
	Understanding information by matching the information	7	2, 3, 5, 9, 12, 17, 18
	Describing the information	3	6, 14, 21
	Evaluating, concluding, and inferencing the text	5	4, 11, 13, 15, 19

The Instrument

*The red sign is the key answers

TEXT 1: Green Area in the City Center: Me and *Wuluh* Starfruit

"I offered *wuluh* starfruit tree seeds for our neighbors. If they interested to join for planting the seeds, our area will be shady. Because now the shady part is not spread evenly, (I am complaining)". "They have several excuses for this issue, they said that it is difficult to get the water and they will wait for the raining season". I am really disappointed, dad".

"Hmmm.. they don't want to use lots of water for watering the trees. Currently, it is better to save the water," daddy said. Yes, sure if it is difficult to get the water, they will not plant the starfruit seeds (I said while scowling).

"Don't worry, the issue of lacking water for planting the trees will be solved with the neighborhood association's project", daddy said. "Our area will have Pandora L"

"Pandora L? What is that"

"This is a mock-up of the wastewater recycling that has been built in our area. It is built in the ground and it is useful for processing household wastewater, such as dishwater" Daddy said.

The water that has been proceeded into the Pandora L can be used for watering the plants and washing cars. The Pandora L installation is located in the Genteng Candirejo, in the center of Surabaya, East Java.

(The text is adapted from *"green area in the city center: me and wuluh starfruit"* by Tyas KW)

Mock-up Pandora L (Waste water recycling)

Question 1
If the neighborhood association's project is successfully held, will the neighbor agree for planting the *wuluh* starfruit?

- ○ No, they will be lazy to plant and take care of *wuluh* starfruit.
- ○ Yes, because Genteng Candirejo has succeeded in exemplifying how to proceed with wastewater.
- ○ No, because they don't want to use lots of water in the dry season.
- ○ Yes, because their worries regarding the lack of water in the dry season will be solved with Pandora L installation.

Question 2
Which one is the discussion topic between Father and I in the text? Give the checklist (√) in every following true statement!

- ▪ Lack of water in the dry season.
- ▪ The equipment for wastewater recycling.
- ▪ Procurement cost for Pandora installation.
- ▪ Planting *wuluh* starfruit.

Question 3
Compare the action of Father and I to react to the situation in the text! Give the checklist (√) in the column of "I", Father", or "I and Father" for every statement.

Statements	I	Father	I and Father
The person who cares the most about their environment			√
The person who has an idea to solve the issue		√	
The person who persuades surrounding to change the environment	√		
The person who is disappointed because the neighbors don't want to plant the trees.	√		
The person who is optimistic the most that Pandora will be a solution to the lack of water in the dry season		√	

Question 4
We can conclude from the text that the conversation between I and Father is held in the dry season

Question 5
Match the statements that are mentioned in the text with the aforementioned equipment!

Statements	
Equipment that I offered to my neighbors	○
Equipment that will be possessed by people in their area	○
Thing that is suggested by the father to be saved	○

Tools	
○	Pandora L (2)
○	Water (3)
○	*Wuluh* starfruit trees
○	*Wuluh* starfruit seeds (1)

Question 6
When I told my father about my experience after talking with our neighbors, I started scowling. Explain the reason why do I express this emotion to my father!
Because I am disappointed with my father's answer. He didn't support my idea of planting the *wuluh* starfruits seeds in our area.

TEXT 2: STEM Practical Activity

Students of SMP Negeri 23 Bandung are sad because of a clean water crisis in their school. According to this reason, they did an experiment to create a simple water-purifying. The water condition in their school from the infiltration wells is muddy, yellowish, and iron-smelly. It is not possible to be used for daily activity, such as make ablution.

A science teacher, Amalia Sholihah engaged students to find the solution to the issue by using research activity. "In the beginning, they felt demotivated because they didn't familiar with the practical and research activities that they need to formulate the problem, search literature, and solve it by themselves. However, after finding the results, they were really happy". Amalia said.

According to the result of their literature review, they discovered affordable and effective materials to purify the water, namely zeolite (a form of gravel with small and medium shapes), active sand, activated charcoal, and aquarium filter. We measured the materials and arranged them in unused places, such as used mineral water bottles or pipes.

From the experiment, the most effective arrangement to purify the water is the small shape of zeolite located in the lowest position, followed by activated charcoal, active sand, a medium shape of zeolite, and an aquarium filter. The results showed that the dirty, smelly, and yellowish water became clean water. The water was not blockage as well.

It is not only effective material but also affordable materials for students. The average price for every material is ranged between 3,000 and 12,000 rupiahs. "If we check from the internet, the price of commercial water purifying is 2,000,000 rupiahs and it is impossible to be bought because the majority of students are from low socioeconomic status", Amalia said.

The experiment is not only useful but also the result of STEM teaching and learning activity of students of SMP Negeri 23 is displayed in several student exhibitions at the level of regional and national. Amalia hopes that students will continue the process by designing the packaging of their water purifying to be sold, so it will give a profit to her students' families.

*STEM (Science Technology Engineering Mathematics) is a term in an academic field for supporting the development of science and technology.

*Make ablution: clean our face, hand, hair, and feet before praying (for Muslims)

(Source: The Article entitled A story of teaching and learning STEM in SMPN 23 Bandung, downloaded from https://edukasi.kompas.com/read/2018/12/20/10462921/sebuah-kisah-praktik-baik-pendidikan-stem-dari-smpn-23-bandung).

Question 7

Who is the main source of information in the text?

- Students
- Teacher
- Head of school
- Parents

Question 8

Mention two effective materials for purifying the water based on the information from the text!

activated charcoal, active sand, zeolite, and an aquarium filter

Question 9

Match the aforementioned paragraph on the left side with the main idea of the paragraph on the right side!

Paragraphs	
1st	○
3rd	○
5th	○

Main ideas	
○	Design an effective and affordable prototype (3)
○	Solve the environmental problem (1)
○	Analyze the economic potential of the developed water purifying (5)
○	Exhibit the water purifying in the student exhibition

Question 10

What is the effect of using a water purifying in the SMP Negeri 23 Bandung? You can choose more than one answer.

- The water will not be yellowish and muddy.
- The developer will be well-known and gets profits.
- It fulfils clean water for making ablution and other activities.
- School can generate sustainable energy.

Question 11

According to the information in the text, the main difference between water purifying developed by students and commercial water purifying is....

- Price
- Quality
- Package
- Effectivity

Question 12

Categorize the following statements into “Fact” or “Opinion” according to the information in the text!

Statements	Fact	Opinion
The quality of school water from the infiltration well is yellowish and smelly	√	
After realizing their research results, all students are happy		√
The materials’ prices for creating water-purifying are cheaper than the commercial water purifying	√	
Commercial water purifying is impossible to be bought by students from low socioeconomic status.		√

Question 13

Why do students are expected to join an activity to solve their environmental problems? You can choose more than one answer.

- Problem-solving is a bridge for the teaching and learning process.
- Environmental problem is the main problem to be solved by students.
- Students have a responsibility to contribute to their environment.
- Solving a problem can be a source to gain a profit.

Question 14

You will find life lessons after reading the text. Write two life lessons that you received from the text!

- We have to care about environmental problems.
- We have to be innovative
- Developing a useful tool is not always expensive
- We can learn a lot while solving an environmental problem.

TEXT 3: 'Strong women' by Hartoyo Andangjaya, 1963

Women who bring small tradeswomen in the early morning,
Where are they come from?
They come from village hills before the train whistle beeps at the railway station
Before people goes to their work.

Women who bring small tradeswomen on the train,
Where will they go?
They ride on steel wheels.
They compete with the sun in the city.
Struggle for life in the city markets.

Women who bring small tradeswomen in the early morning,
Who are they?
They are strong mothers and wonder women.
Roots creep from the ground of the hills down to the city.
They show their love by supporting their family from one village to another village.

(Hartoyo Andangjaya, 1963)
(downloaded from https://indonesianliteraryworks.blogspot.com/2016/09/poetry-perempuan-perempuan-perkasa.html)

Question 15
What is your conclusion regarding women in the poem?
Hardworking and strong women.

Question 16
Where is the location described in the first paragraph of the poem?

- ○ Railway station and village
- ○ Village hill and city market
- ○ City market and railway station
- ○ Railway station and city

Question 17
Match the paragraph in the poem from the left side with the main idea on the right side!

Paragraph	
1st	○
2nd	○
3rd	○

Main idea	
○	Telling about who are the strong women (3)
○	Telling about where the strong women come from (1)
○	Telling about where will the strong women go (2)
○	Telling about how the strong women sell their stuff

Question 18

Are these statements consistent with the information in the poem?

Statements	Yes	No
The strong women are sellers from the village who want to sell their stuffs in the city	√	
The strong women come from the village hills to the railway station at night		√
The strong women are together going to the city	√	
The strong women are fighting for spaces in the city market		√

Question 19

After reading the poem, what personalities are possible to be imitated? You can choose more than one answer

- Hardworking
- Frugal
- Diligent
- Lazy

Question 20

Who are the strong women in the poem?

The women who come from the village to sell their stuffs in the city

Question 21

What is their aim to bring the small tradeswomen on the train?

To bring their stuffs to be sold to the city

The Construct of the Scoring Rubric

Type of questions	Items	Score
Multiple-choice	1, 7, 11, 16	1: correct, 0: incorrect
Matching and true/false	2, 3, 5, 9, 10, 12, 13, 17-19	Depending on the statements (suppose the statements to be matched are 4, then the score is ranged between 0 and 4) Items 2, 5, 9, and 17: 0-3 Item 3: 0-5 Items 10, 13, and 19: 0-2 Items 12, and 18: 0-4
Short answer	4, 8, 15	Depending on the question, if the questions asked to mention an answer, then the maximum score is 1. If the questions asked to mention two answers, then the maximum score is 2. Item 4: 0-1 Item 8 and 15: 0-2
Long answer	6, 14, 20, 21	0-5 0: blank answer 1: irrelevant answer 2: incomplete and major problem 3: complete and major problem 4: minor problem 5: correct answer

TEXT COMPREHENSION TEST
(Original Indonesian Version)

Nama :
Kelas :
Jenis Kelamin :
Sekolah :

Kerjakan soal dibawah ini dengan seksama!

Hijau Kampungku di Tengah Kota: Aku dan Belimbing Wuluh

“Aku tadi menawarkan bibit pohon belimbing kepada beberapa tetangga. Kalau mereka ikut menanam pohon, jalan kampung ini akan menjadi lebih teduh. Tidak seperti sekarang, ada bagian yang teduh, ada bagian yang panas,” keluhku. “Ada yang beralasan, air sedang susah didapat. Ada yang berkata, 'Nanti, ya, tunggu musim hujan datang lagi.' Aku kecewa, Yah.”
“Hmm... Mereka tidak ingin menggunakan air terlalu banyak. Saat ini, memang sebaiknya kita hemat air,” kata Ayah. Tentu saja, aku semakin merengut. Kalau air tetap sukar didapat, tidak ada orang yang mau menanam pohon belimbing wuluh itu.
“Tapi, jangan khawatir. Masalah air untuk menyiram tanaman sebentar lagi akan terbantu oleh proyek Pak RT,” kata Ayah. “Kampung kita akan punya Pandora L.”
“Pandora L? Apa itu, Yah?”

Maket Pandora L (Pendaur Ulang Limbah Air)

“Ini maket dari pengolahan limbah yang baru saja selesai dibangun di kampung kita. Bangunan ini ada di dalam tanah dan berguna untuk mengolah limbah rumah tangga saja, seperti air cucian,” kata Ayah.
Air hasil pengolahan dari Pandora L digunakan untuk kegiatan menyiram tanaman dan mencuci kendaraan. Pemasangan Pandora L terletak di Kampung Genteng Candirejo di tengah kota Surabaya, ibu kota Provinsi Jawa Timur.
(Diadaptasi dari *Hijau Kampungku di Tengah Kota: Aku dan Belimbing Wuluh* karya Tyas KW)

Soal 1

Jika proyek Pak RT telah dilaksanakan, apakah para tetangga akan setuju untuk menanam pohon belimbing wuluh?

- o Tidak, warga akan tetap saja malas menanam dan merawat pohon belimbing wuluh.
- o Ya, karena Kampung Genteng Candirejo telah berhasil mencontohkan cara pengelolaan air limbah.
- o Tidak, karena warga tidak ingin menggunakan air terlalu banyak saat musim kemarau.
- o Ya, karena kekhawatiran kurangnya air di musim kemarau akan terpecahkan dengan dipasangnya Pandora.

Soal 2

Manakah hal-hal yang menjadi topik pembicaraan antara tokoh Aku dan Ayah? Beri tanda centang (√) pada setiap pernyataan yang benar!

- ▪ Kekurangan air di musim kemarau.
- ▪ Alat untuk mengolah air.
- ▪ Pengadaan biaya untuk pemasangan Pandora.
- ▪ Penanaman belimbing wuluh.

Soal 3

Bandingkan sikap tokoh Aku dan Ayah terhadap situasi yang ada pada cerita tersebut! Beri tanda centang (√) pada kolom “Aku”, “Ayah”, atau “Aku dan Ayah” untuk setiap pertanyaan.

Pernyataan	Aku	Ayah	Aku dan Ayah
Tokoh yang peduli lingkungan di tempat tinggal mereka			√
Tokoh yang memiliki ide untuk mnegatasi masalah yang dihadapi		√	
Tokoh yang mengajak orang-orang disekitar untuk melakukan Gerakan perbaikan	√		
Tokoh yang kecewa karena para tetangga enggan diajak menanam pohon	√		
Tokoh yang optimis bahwa Pandora akan menjadi solusi kurangnya air di musim kemarau		√	

Soal 4

Dari wacana tersebut, dapat disimpulkan bahwa perbincangan antara tokoh Aku dan Ayah terjadi pada musim

…………………………………………………………………………………………………

Soal 5

Pasangkanlah sesuatu yang disebutkan oleh tokoh dalam cerita tersebut dengan benda yang dimaksud!

Pernyataan	
Hal yang ditawarkan oleh tokoh Aku kepada tetangganya	o
Hal yang akan segera dimiliki oleh warga kampung	o
Hal yang disarankan oleh tokoh Ayah untuk dihemat	o

	Benda
o	Pandora L
o	Air
o	Pohon belimbing wuluh
o	Bibit pohon belimbing (1)

Soal 6
Ketika tokoh Aku menceritakan kejadian yang baru saja dialaminya kepada Ayah, tokoh Aku justru merengut. Jelaskan alasan tokoh aku menunjukkan eskpresi tersebut kepada Ayah!
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

Praktik STEM

Berangkat dari keprihatinan siswa SMP Negeri 23 Bandung melihat krisis air bersih di sekolah, mereka bereksperimen membuat alat penjernih air sederhana. Kondisi air di sekolah yang bersumber dari air sumur resapan warnanya kuning dan keruh, serta berbau besi, tentu saja tidak dapat dipergunakan untuk aktivitas sehari-hari, seperti wudu dan buang air.

Guru pembimbing, Amalia Sholihah menangkap keprihatinan para siswa didiknya kemudian mengajak mereka mencari solusi dengan melakukan riset mandiri. "Ketika membuat ini anak-anak sempat stres, karena tidak terbiasa. Biasanya kan berupa resep, kalau ini harus menggali, mencari tahu sendiri, tapi begitu lihat hasilnya mereka sangat bahagia," tutur Amalia guru IPA.

Dari hasil penelitian, siswa menemukan bahan-bahan yang harganya terjangkau tapi efektif menjernihkan air, yaitu ziolit berbentuk seperti kerikil dengan ukuran kecil dan sedang, pasir aktif, arang aktif, dan filter akuarium. Bahan-bahan ini kemudian ditakar dan disusun pada wadah yang sudah tidak terpakai, seperti botol air mineral bekas atau pipa.

Dari percobaan yang dilakukan, susunan paling efektif untuk menjernihkan air adalah ziolit dengan ukuran kecil pada posisi paling bawah, dilanjutkan arang aktif, pasir aktif, lalu diisi kembali dengan ziolit berukuran sedang. Terakhir, posisi teratas dipasang filter akuarium. Hasilnya, ketika air tercemar dituang, air yang semula kuning, keruh, dan berbau, menjadi bening dan tidak berbau sama sekali. Air juga dapat mengalir dengan lancar, tidak mengalami penyumbatan.

Bukan sekadar efektif, namun bahan-bahan tersebut harganya pun terjangkau, sehingga terbeli oleh siswa. Masing-masing bahan tersebut harganya berkisar antara tiga ribu hingga dua belas ribu rupiah. "Kalau kita lihat di internet harga filter itu dua juta, tidak mungkin terbeli oleh anak-anak saya yang keluarganya menengah ke bawah," ungkap Amalia.

Selain dapat dirasakan langsung manfaatnya, hasil pembelajaran STEM siswa SMP Negeri 23 Bandung ini juga seringkali diikutkan pada ekspos karya pelajar, baik di tingkat kota, provinsi, maupun nasional. Tidak puas hanya sampai di sini, Amalia ingin para siswa dapat mengemas penjernih air dalam wadah menarik, sehingga memiliki nilai ekonomis. "Lumayan untuk pemasukan, membantu ekonomi keluarga mereka," harap Amalia.

*STEM (Science Technology Engineering Mathematics) atau sains, teknologi, teknik, dan matematika merupakan istilah yang dipakai dalam sekelompok pelajaran akademik untuk menunjang pengembangan sains dan teknologi.

*wudu: menyucikan diri (sebelum salat) dengan membasuh muka, tangan, kepala, dan kaki.

(Sumber: Artikel berjudul Sebuah Kisah Praktik Baik Pendidikan STEM dari SMPN 23 Bandung, diunduh dari https://edukasi.kompas.com/read/2018/12/20/10462921/sebuah-kisah-praktik-baik-

pendidikan-stem-dari-smpn-23-bandung).

Soal 7
Siapakah yang memaparkan informasi di dalam teks tersebut?

- o Siswa
- o Guru
- o Kepala Sekolah
- o Orang tua

Soal 8
Sebutkan dua bahan yang efektif menjernihkan air sesuai dengan informasi yang tersedia di dalam teks!
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………..

Soal 9
Pasangkanlah urutan paragraf yang ada di lajur kiri dengan pokok pembahasannya yang ada di lajur kanan!

Urutan Paragraf	
Pertama	o
Ketiga	o
Kelima	o

Benda	
o	Merancang karya yang efektif dan terjangkau
o	Memecahkan masalah lingkungan sekitar
o	Melihat potensi ekonomis dari karya yang dibuat
o	Menampilkan karya di pameran pelajar

Soal 10
Apa saja dampak yang akan terjadi jika karya penjernih air dapat digunakan di lingkungan sekolah SMP Negeri 23 Bandung? Kamu bisa memilih lebih dari satu jawaban.

- o Air sekolah tidak lagi berwarna kuning dan keruh.
- o Pencipta karya akan terkenal dan mendapatkan banyak keuntungan.
- o Fasilitas wudu dan buang air dapat menyediakan air yang lebih bersih.
- o Sekolah dapat menghasilkan energi yang ramah lingkungan.

Soal 11
Menurut informasi yang ada di dalam teks, perbedaan utama antara alat penjernih air buatan siswa dan alat filter air yang umumnya dijual di internet terkait dengan ….

- o Harga
- o Kualitas
- o Kemasan
- o Efektivitas

Soal 12
Kategorikanlah pernyataan berikut ke dalam “Fakta” atau “Opini” sesuai dengan informasi yang ada di dalam teks!

Pernyataan	Fakta	Opini
Kualitas air sekolah yang bersumber dari sumur resapan berwarna kuning dan berbau besi	√	
Begitu melihat hasil karyanya semua murid sangat bahagia		√
Bahan-bahan untuk alat penjernihan air harganya lebih murah dibandingkan filter air di internet	√	
Filter air berharga jutaan tidak mungkin terbeli oleh siswa yang berasal dari keluarga menengah kebawah		√

Soal 13
Mengapa murid sekolah diharapkan dapat ikut memecahkan permasalahan yang ada di lingkungan sekitar? Kamu bisa memilih lebih dari satu jawaban.

- o Memecahkan permasalahan dapat menjadi sarana kegiatan untuk pembelajaran.
- o Lingkungan sekitar adalah masalah utama yang harus segera dicari solusinya oleh seorang murid.
- o Pada dasarnya murid memiliki tanggung jawab untuk ikut berkontribusi positif terhadap lingkungannya.
- o Memecahkan permasalahan dapat menjadi sumber penghasilan tambahan untuk mencari keuntungan.

Soal 14
Setelah membaca teks informasi tersebut, tentunya kamu memperoleh pesan yang dapat kamu terapkan dalam kehidupan sehari-hari.
Tuliskan dua pesan yang kamu peroleh dari teks informasi tersebut!
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………..

'Perempuan – perempuan perkasa' karya Hartoyo Andangjaya, 1963

Perempuan-perempuan yang membawa bakul di pagi buta,
dari manakah mereka..
Ke stasiun kereta mereka datang dari bukit-bukit desa sebelum peluit kereta pagi terjaga..
Sebelum hari bermula dalam pesta kerja..

Perempuan-perempuan yang membawa bakul dalam kereta,
ke manakah mereka..
Di atas roda-roda baja mereka berkendara.
Mereka berlomba dengan surya menuju gerbang kota..
Merebut hidup di pasar-pasar kota..

Perempuan-perempuan perkasa yang membawa bakul di pagi buta,
siapakah mereka..
Mereka ialah ibu-ibu berhati baja, perempuan-perempuan perkasa.
akar-akar yang melata dari tanah perbukitan turun ke kota..
Mereka cinta kasih yang bergerak menghidupi desa demi desa..

(Hartoyo Andangjaya, 1963)
(diunduh dari https://indonesianliteraryworks.blogspot.com/2016/09/poetry-perempuan-perempuan-perkasa.html)

Soal 15
Simpulan apa yang kamu dapat mengenai perempuan dalam puisi tersebut?
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………..

Soal 16
Latar tempat yang disampaikan di dalam bait pertama puisi adalah ….
- o Stasiun dan bukit desa
- o Bukit desa dan pasar kota
- o Pasar kota dan stasiun
- o Stasiun dan gerbang kota

Soal 17
Pasangkanlah bait puisi yang ada di lajur kiri dengan isis puisi yang ada di lajur kanan!

Bait puisi	
Pertama	o
Kedua	o
Ketiga	o

Isi puisi	
o	Tentang siapakah yang dimaksud dengan perempuan-perempuan perkasa (3)
o	Tentang darimanakah perempuan-perempuan perkasa berasal (1)
o	Tentang ke manakah perempuan-perempuan perkasa akan pergi (2)
o	Tentang bagaimanakah perempuan-perempuan perkasa berdagang

Soal 18
Apakah pernyataan-pernyataan berikut sesuai dengan informasi yang ada di dalam puisi?

Pernyataan	Sesuai	Tidak sesuai
Perempuan-perempuan perkasa adalah ibu-ibu pedagang dari desa yang berjualan ke kota	√	
Perempuan-perempuan perkasa berangkat dari perbukitan desa menuju ke stasiun di malam hari		√
Perempuan-perempuan perkasa berangkat beramai-ramai menuju ke kota	√	
Perempuan-perempuan perkasa saling berebut tempat di pasar kota		√

Soal 19
Jika kamu membaca puisi tersebut, sikap apa yang dapat ditiru dari tokoh dalam puisi tersebut?
- o Kerja keras
- o Hemat
- o Rajin
- o Santai

Soal 20

Siapakah perempuan-perempuan perkasa yang dimaksud oleh pengarang pada puisi tersebut?

……
……
……..

Soal 21

Untuk apa perempuan-perempuan itu membawa bakul ke kereta?

……
……
……..

APPENDIX D

SCIENCE KNOWLEDGE TEST

The Construct of the Test

Format	Indicators	Items		
		Grades 7	Grades 8	Grades 9
Photosynthesis				
Multiple-choice	Understand the organs of photosynthesis	1	1	1
	Understand the role of water in photosynthesis and how it can be absorbed	2	2	2
	Understand the experiments related to photosynthesis	3	3	3
	Understand the factors affecting photosynthesis	4	4	4
	Understand the outputs or results from photosynthesis	5, 6	5, 6	5, 6
	Understand the role of sun in photosynthesis	7	7	7
	Understand the definition, role and function of chlorophyll	8	8	8
	Understand the relation of photosynthesis with other things (respiration, technology)	9	9	9
Energy, Nutrition, and health				
Multiple-choice	Mention the example of food that contains nutrients	6, 7, 9	3, 6, 8, 9	3, 8
	Understand the function of nutrients	1, 4, 5, 8	2, 4, 5, 7	4—7, 9
	Understand the energy balanced	-	-	1
	Understand experiments related to energy	2, 3	-	-
	Mention the factors affect human energy	-	1	2
Biotechnology				
Multiple-choice	Understand the field applied in biotechnology	-	1	1
	Mention the examples of biotechnology products	1, 9	2, 5	2
	Understand the applications of biotechnology	2, 3, 7	6, 9	6, 7
	Understand the foundations of biotechnology (conventional & modern)	5, 6	3, 4	3, 4
	Understand the effects of biotechnology	8	7	8, 9
	Understand the used of other living things in biotechnology	4	8	5
Eco-friendly Technology				
Multiple-choice	Understand the examples of eco-friendly technology and their application	1, 6, 7, 9	1, 5, 6, 8	1, 3, 4, 5, 7
	Understand the negative effects of non-eco-friendly technology products	3	9	8
	Understand the definition of eco-friendly technology	4	-	-
	Understand the principle of eco-friendly technology	2, 5	2, 3	2, 9
	Understand the advantages and disadvantages of using eco-friendly technology	8	7, 4	6

The Construct of the Test (cont.)

Format	Indicators	Items		
		Grades 7	Grades 8	Grades 9
Climate Changes				
Multiple-choice	Understand the impacts or effects of global warming	1, 2, 3, 8	5, 7, 8	1, 2, 3
	Understand the preventive actions of global warming	4	4, 6	8
	Understand the activities and gases that caused global warming	5	3, 2	5, 6, 7, 9
	Mention the location of the ozone	6	-	-
	Understand the function of ozone	7	-	-
	Understand the preventive actions of depletion of ozone	9	9	-
	Mention the eco-friendly fuel	-	1	4
Scientific Methods				
Multiple-choice	Understand the steps of the scientific method	1, 4, 5, 8	4	1, 5
	Understand the definition of the scientific method and the definition of every step of the scientific method	2, 7	7, 3	4, 9
	Understand the variables in the scientific research	3	5	-
	Understand the scientific attitude and the characteristics of the scientific method	6	1, 9	6
	Understand the type of data	9	6, 8	3, 7, 8
	Understand the benefits of the scientific method	-	2	-
	Understand the function of tools in experiment	-	-	2

APPENDIX E

CHECKLIST VALIDATION INSTRUMENTS

Instruments	Validity	Reliability
Integrated STEM-based mathematical problem-solving test	• Content validity (experts) • Construct validity (Rasch analysis)	• Internal reliability (EAP/PV reliability) • Inter-rater reliability (ICC)
Mathematics DSPK test	• Content validity • Construct validity	• Internal reliability
Text comprehension test	• Construct validity	• Internal reliability
Science knowledge test	• Content validity	• Internal reliability

APPENDIX F

ETHICAL APPROVAL

University of Szeged

Institutional Review Board
Doctoral School of Education

6722 Szeged, 30-34 Petőfi S. Av., Hungary
Phone/fax: +36 62 544-032

Ijtihadi Kamilia Amalina
PhD Student: Doctoral School of Education
Reference number: 7/2022
Subject: Ethical evaluation of a research project

Date: 5 July, 2022

E T H I C A L A P P R O V A L

The Insitutional Review Board (IRB) of the Doctoral School of Education, University of Szeged has recently reviewed your application for an ethical approval (Title of the Research Project: "**Assessment of Students' Integrated STEM based Mathematical Problem-Solving Skills and the Factors Influencing them**", supervisor: Dr. Tibor Vidákovich). This proposal is deemed to meet the requirements of the ethical conducts on social research with human subjects of the Doctoral School of Education, University of Szeged.

IRB decision: approved

Justification:
The research project meets the requirements of the professional-ethical criteria of the social research including human subjects within the field of education science. Main goal of the study is to investigate the development of students' integrated STEM based mathematical problem-solving skills across grades cohort and to examine the impact of a social-emotional factor on integrated STEM based mathematical problem-solving skills. The participants will be is 7th-9th graders junior secondary school students in East Java, Indonesia (N = approx. 600-900) ages from 12-15 years. Computer-based surveys, tests and essays will be applied as methods of data collection, replies will be registered by name/code. The original data could only be seen by leader of research. The school will request consent from the parents. The participation is voluntary. Procedure of the data collection does not harm their privacy law, it does not have an impact on the participants' mental or physical health. Data cannot be handled by persons to whom they are not concerned.

In a summary, full ethical approval has been granted.

We wish you all the best for the conduct of the project.

Prof. Dr. Bettina Pikó
IRB coordinator