

**UNIVERSITY OF SZEGED**  
**DOCTORAL SCHOOL OF EDUCATION**



**THE ASSESSMENT OF DOMAIN-GENERAL AND DOMAIN-SPECIFIC PROBLEM-  
SOLVING IN INDONESIA: EXPLORING THE ROLE OF INDUCTIVE  
REASONING, SCIENTIFIC COMPETENCY AND AFFECTIVE FACTORS**

**PHD DISSERTATION**  
**BY**  
**AZIZUL GHOFAR CANDRA WICAKSONO**

**SUPERVISOR: DR. ERZSÉBET KOROM**

**SZEGED, HUNGARY**

**2024**

## TABLE OF CONTENT

<b>LIST OF TABLES</b> .....	5
<b>LIST OF FIGURES</b> .....	6
<b>ABBREVIATIONS</b> .....	8
<b>CHAPTER 1 INTRODUCTION</b> .....	10
<b>1.1 Background of the study</b> .....	10
<b>1.2 Research context</b> .....	11
<b>1.3 The structure of the dissertation</b> .....	13
<b>CHAPTER 2 LITERATURE REVIEW</b> .....	14
<b>2.1 The concept of problem-solving</b> .....	14
<b>2.1.1 Definition of a problem and solving a problem</b> .....	14
<b>2.1.2 Domain-general and domain-specific problem-solving</b> .....	15
<b>2.1.3 The assessment of domain-general and domain-specific problem-solving</b> .....	17
<b>2.1.4 The assessment of problem solving in Indonesia</b> .....	26
<b>2.1.5 The role of gender and socioeconomic status in problem solving</b> .....	34
<b>2.2 The cognitive factors affecting problem solving</b> .....	35
<b>2.2.1 Inductive reasoning</b> .....	35
<b>2.2.2 The role of inductive reasoning in problem solving</b> .....	38
<b>2.2.3 Role of gender in inductive reasoning</b> .....	39
<b>2.2.4 Scientific competencies and knowledge</b> .....	40
<b>2.2.5 The role of scientific competencies and knowledge in problem solving</b> .....	43
<b>2.2.6 Role of gender in scientific competency</b> .....	44
<b>2.3 The affective factors influencing problem solving</b> .....	45
<b>2.3.1 Attitude towards science</b> .....	45
<b>2.3.2 The role of attitudes in problem solving</b> .....	48
<b>2.3.3 Gender differences in attitude</b> .....	49
<b>2.3.4 Motivation in science</b> .....	50
<b>2.3.5 The role of motivation in problem solving</b> .....	52
<b>2.3.6 Gender differences in motivation</b> .....	54
<b>2.4 The theoretical model of the connection between cognitive and affective factors, gender, and socioeconomic status on problem solving</b> .....	54
<b>CHAPTER 3 RESEARCH AIMS, QUESTIONS, AND HYPOTHESES</b> .....	57
<b>3.1 Aims of the research</b> .....	57
<b>3.2 Research questions and hypotheses</b> .....	57
<b>CHAPTER 4 RESEARCH METHODOLOGY</b> .....	59

4.1 Research design .....	59
4.2 Sample .....	59
4.3 Instruments .....	60
4.3.1 Domain-general problem-solving test: MicroDYN.....	60
4.3.2 Domain-specific problem-solving test.....	62
4.3.3 Inductive reasoning test .....	64
4.3.4 Scientific competency test .....	65
4.3.5 Attitudes towards science questionnaire.....	66
4.3.6 Science motivation questionnaire.....	67
4.3.7 Background questionnaire.....	68
4.4 Procedures.....	68
4.5 Data analysis .....	68
<b>CHAPTER 5 EMPIRICAL RESULTS.....</b>	<b>70</b>
<b>5.1 The assessment of domain-general problem-solving.....</b>	<b>70</b>
5.1.1 The validation of MicroDYN test .....	70
5.1.2 The ability level of the student's domain-general problem-solving .....	73
5.1.3 The gender difference in domain-general problem-solving ability .....	74
5.1.4 Discussion .....	76
<b>5.2 The assessment of domain-specific problem-solving.....</b>	<b>78</b>
5.2.1 The validation of science problem-solving test .....	78
5.2.2 The ability level of students' domain-specific problem-solving.....	80
5.2.3 The gender difference in domain-specific problem-solving ability .....	81
5.2.4 Discussion .....	81
<b>5.3 The assessment of inductive reasoning .....</b>	<b>84</b>
5.3.1 The validation of inductive reasoning test.....	84
5.3.2 The ability level of student's inductive reasoning ability .....	87
5.3.3 The gender difference in inductive reasoning ability .....	87
5.3.4 Discussion .....	88
<b>5.4 The assessment of scientific competency .....</b>	<b>90</b>
5.4.1 The validation of scientific competency test.....	91
5.4.2 The ability level of students' scientific competency .....	93
5.4.3 The gender difference in scientific competency .....	94
5.4.4 Discussion .....	95
<b>5.5 The assessment of attitudes towards science.....</b>	<b>97</b>
5.5.1 The validation of attitude towards science questionnaire.....	97
5.5.2 The profile of students' attitude towards science.....	100

5.5.3 The gender difference in attitude towards science .....	101
5.5.4 Discussion .....	102
5.6 The assessment of science motivation .....	104
5.6.1 The validation of Science Motivation Questionnaire.....	105
5.6.2 The profile of students' science motivation .....	107
5.6.3 The gender difference in science motivation .....	108
5.6.4 Discussion .....	109
5.7 The relationship between cognitive and noncognitive factors on problem solving .....	111
5.7.1 The result of empirical analysis.....	112
5.7.2 Discussion .....	114
<b>CHAPTER 6 GENERAL CONCLUSION, IMPLICATION, LIMITATION, AND RECOMMENDATION.....</b>	<b>119</b>
6.1 General conclusion .....	119
6.2 Educational implication .....	122
6.3 Limitation and future recommendation .....	123
<b>REFERENCES .....</b>	<b>125</b>
<b>APPENDIX A. MicroDYN test .....</b>	<b>148</b>
<b>APPENDIX B. Science Problem-solving test .....</b>	<b>153</b>
<b>APPENDIX C. Inductive Reasoning test .....</b>	<b>160</b>
<b>APPENDIX D. Scientific Competency test .....</b>	<b>169</b>
<b>APPENDIX E. Attitude Towards Science Questionnaire .....</b>	<b>176</b>
<b>APPENDIX F. Science Motivation Questionnaire II .....</b>	<b>177</b>

## LIST OF TABLES

Table 1. The description of the creative problem-solving framework in PISA 2012.....	25
Table 2. The problem-solving test developed in Indonesia.....	28
Table 3. The problem-solving framework used in the literature study.....	29
Table 4. Topic distribution of problem-solving assessment tools in Indonesia.....	30
Table 5. The types of inductive reasoning process.....	37
Table 6. The context of PISA science assessment.....	42
Table 7. The characteristics of the participants.....	60
Table 8. The complexity of the MicroDYN test.....	62
Table 9. The description context of science problem-solving test.....	63
Table 10. Description of the scientific competency test.....	66
Table 11. The item fit of the MicroDYN test from Rasch analysis.....	71
Table 12. The gender comparison of students' domain-general problem-solving.....	75
Table 13. The item fit of science problem-solving test.....	79
Table 14. The gender comparison of students' science problem-solving.....	81
Table 15. The item fit of the inductive reasoning test.....	85
Table 16. The gender comparison of students' inductive reasoning ability.....	88
Table 17. The item fit analysis results for scientific competency.....	91
Table 18. The gender comparison of students' scientific competency.....	95
Table 19. The item fit analysis results for the ATSQ.....	98
Table 20. The gender comparison of students' attitude towards science.....	102
Table 21. The item fit analysis results for the SMQ II.....	105
Table 22. The gender comparison of students' science motivation.....	109
Table 23. The correlation between gender, socio-economic status, cognitive and affective factors, and problem solving.....	112
Table 24. Standardised results of model analysis.....	114

## LIST OF FIGURES

Figure 1. The example of Taylorshop task .....	18
Figure 2. The example of multiflux task .....	19
Figure 3. The scheme of variable involved in the MicroDYN test .....	20
Figure 4. Screen capture of the MicroDYN task in “gas and vehicle” during the application phase.....	21
Figure 5. The items sample of the biology problem solving test .....	23
Figure 6. The items sample of the science problem solving test in high school .....	24
Figure 7. The example of PISA problem solving test. ....	26
Figure 8. The heuristic-analytic theory of reasoning .....	36
Figure 9. Interrelation between scientific competency components .....	42
Figure 10. The framework of three dimensions of attitude towards science .....	48
Figure 11. The theoretical model of the relationship between cognitive and affective factors, gender, and socioeconomic status to problem solving .....	56
Figure 12. The knowledge exploration phase. ....	61
Figure 13. The knowledge acquisition phase. ....	62
Figure 14. An example of a science-problem solving test in the generating solution phase. ..	64
Figure 15. The inductive reasoning tasks .....	65
Figure 16. The example of scientific competency test.....	66
Figure 17. The Wright map of the MicroDYN test .....	72
Figure 18. The DIF results for MicroDYN test.....	73
Figure 19. The distribution of students’ logit scores in the MicroDYN test .....	74
Figure 20. The distribution of students’ problem-solving scores in tasks of different complexity.....	74
Figure 21. The distribution of students’ domain-general problem-solving.....	75
Figure 22. The Wright map of the science problem-solving test .....	79
Figure 23. The DIF result for science problem-solving test items.....	80
Figure 24. The distribution of students’ logit scores in science problem solving .....	80
Figure 25. The distribution of students’ science problem-solving ability .....	81
Figure 26. The Wright map of the inductive reasoning test .....	86
Figure 27. The DIF result for the inductive reasoning test .....	86
Figure 28. The distribution of students’ logit scores in inductive reasoning .....	87
Figure 29. The distribution of students’ inductive reasoning ability.....	88
Figure 30. The Wright map of the scientific competency test items.....	92
Figure 31. The DIF result of the scientific competency test .....	93
Figure 32. The distribution of students logit score in scientific competency.....	94
Figure 33. The distribution of students’ scientific competency .....	94
Figure 34. The Wright map of the ATSQ items.....	99
Figure 35. The DIF results for ATSQ.....	100

Figure 36. The distribution of students' logit scores in attitude towards science .....	101
Figure 37. The distribution of students' attitude towards science .....	102
Figure 38. The Wright map of the SMQ II items .....	106
Figure 39. The DIF results for SMQ II .....	107
Figure 40. The distribution of students' logit scores in science motivation .....	108
Figure 41. The distribution of students' science motivation .....	109
Figure 42. The relationship model between cognitive and noncognitive factors on problem solving .....	113

## ABBREVIATIONS

ANX	Anxiety and difficulty
ATOSS	Attitude toward organise science scale
ATSQ	Attitude toward science questionnaire
BRAINS	Behaviours, related attitudes, and intentions toward science
BSNP	Badan standar nasional pendidikan
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CM	Career motivation
CPS	Complex problem-solving
DIF	Different item functioning
ENJ	Enjoyment
FA	Figure analogy
FS	Figure series
GM	Grade motivation
GS	Generating solution
IM	Intrinsic motivation
IP	Identifying problem
IR	Inductive reasoning
MCS	Multiple complex system
ML	Maximum likelihood
MNSQ	Mean square
NA	Number analogy
NS	Number series
OECD	Organisation for economic co-operation and development
PAR	Participation in science learning and activities
PISA	Programme for international student assessment
PTMA	Point-biserial correlation
RMSEA	Root mean square error of the approximation
SC	Scientific competency
SCS	Single complex system
SCE	Explaining scientific phenomena
SCI	Interpreting and evaluating scientific data or evidence
SD	Self determination
SE	Self-efficacy
SEM	Structural equation model

SES	Socioeconomic status
SMQ II	Science motivation questionnaire II
SPS	Science problem solving
SRMR	Standardised root mean square residual
STEM	Science, technology, engineering and mathematics
SVAL	Value of science
TLI	Tucker-Lewis index
VOTAT	Vary one thing at time

# CHAPTER 1 INTRODUCTION

## 1.1 Background of the study

Digitalisation and technological development have raised concerns about how people can adapt the dynamic change in their social and daily lives. Those who can adapt to these changes will benefit their future work. However, individuals who lack adequate preparation may face insecurity and vulnerability in the workplace and social environment (OECD, 2017c). To address this challenge, schools must prepare students with skills and abilities for changes that are more rapid than before. Dealing with changes and unfamiliar situations is within the scope of problem-solving construct (Greiff et al., 2012). Problem solving refers to the cognitive process in transforming a given situation into a solved situation when there is no obvious method or solution available (Mayer & Wittrock, 2006). It includes the willingness to engage with such a situation to understand and resolve the solutions (OECD, 2014). Problem solving has been studied for decades and is still considered an important skill in the 21<sup>st</sup> century (Baker & Mayer, 1999; Care et al., 2016; Csapó & Funke, 2017; Dindar et al., 2022; Wirth & Klieme, 2003). Problem solving is one of the basic requirements to work in society which causes to an increasing demand for people with high problem-solving ability in the workplace (Zhou et al., 2012). Indeed, teaching problem solving on an educational level becomes a compulsion to prepare students for their future life.

The integration of problem solving in education has been applied in curricula, for example, the National Research Council (2012) specified problem solving as one of the core skills in K-12 science education in the United States. The Indonesian national curriculum also included problem solving as one of the core competencies to attain secondary education (BSNP, 2013). The integration of problem solving as a core component in curricula has led to a surge in research in the area of learning and instruction. Many studies have implemented a learning strategy based on problem-solving activity in the classroom (Ferreira & Trudel, 2012; Kuo et al., 2012; Lee-Cultura et al., 2022; Liu & Israel, 2022). In addition, the research of problem solving is developing rapidly in the area of educational assessment with the standardised measurement of problem solving developed in general context such as MicroDYN, Tailorshop, MultiFlux, and the Creative and Collaborative problem-solving test (Greiff et al., 2012; OECD, 2014, 2017c; Stadler et al., 2015). Further evaluation of problem solving also applied to specific problem phenomena in the area of science, mathematics, and technology (Bátor & Szeberényi, 2021; Li et al., 2020; Seifried et al., 2020). This type of assessment is used mainly to support the learning process during classroom practices.

However, since problem solving is a complex process, it requires a dynamic connection to both cognitive and affective factors. Problem solving in a specific domain is strongly connected to prior knowledge to understand the problem situation, while general problem-solving does not necessarily require prior knowledge, but is related to intelligence and reasoning processes (Greiff & Neubert, 2014). Furthermore, affective factors such as attitude and motivation foster engagement and the use of effective strategies in problem solving. Thus, including these factors in the assessment process and analysing their connection to problem solving will give comprehensive information to explain the dynamic of problem solving and its influencing factors.

## **1.2 Research context**

The assessment of problem solving has been widely studied in recent years. In domain-general problem-solving, the assessment process is focused on the ability to solve real-world complex problem scenario. Subsequently, measurement tools have been developed to address problem solving based on a complex scenario known as complex problem solving (CPS). CPS is defined as the successful interaction between dynamic environments which involves the successful exploration and integration of information in the problem-solving process. The use of complex problem solving as a general measurement of problem solving has been presented in many studies (Greiff, 2012; Greiff et al., 2015; Stadler et al., 2015; Kyllonen & Kell, 2017). The PISA survey also adapted the CPS framework in global measurement of creative problem solving (OECD, 2014). In addition to general problem-solving, domain-specific problem-solving measurement has been developed since the discovery of mathematical problem solving (Pólya, 1945). The assessment of domain-specific problem-solving is used in connection with learning process, assisting students in problem-solving activities. To this point, the assessment of domain-specific problem-solving has been expanding into various contexts such as mathematics, science, and technology (Bátor & Szeberényi, 2021; Gallagher et al., 2000; Priemer et al., 2020; Szeberényi, 2013). Both domain-general and domain-specific problem-solving offer different benefits in supporting individuals to deal with complex and general or specific problem situation.

Since PISA (OECD, 2014, 2019c) used problem solving as one of its global assessment frameworks, information on student-level performance was collected in many countries to support the further development of their educational system. However, Indonesia did not participate in the problem-solving assessment, leaving no comprehensive information on the level of problem-solving ability of their students compared to other country. As regards the

assessment of problem solving in Indonesia, Wicaksono & Korom (2022) conducted a review on research in problem-solving assessment within Indonesian context and found a limited number of research paper explains the standardised assessment of problem solving. They found all tests addressed problem solving in a specific domain of science and mathematics with less information on the psychometric properties of the tests. Additionally, in the Indonesian educational system, there are national assessments conducted at the end of the school level. These computer-based assessments comprehensively examine students' academic achievement throughout their educational program, focusing on competency and literacy. This condition contradicts the objective of Indonesian curricula that includes problem-solving ability as one of the main focus of their educational purposes (BSNP, 2013). Indeed, there is no recent study of comprehensive problem-solving assessment in the Indonesian context, which makes it important to conduct a study on domain-general and domain-specific problem-solving assessment.

The work on the evaluation of domain-general and domain-specific problem-solving is not enough in itself to explain the dynamics of problem solving in the educational domain. To develop a broader picture of the dynamics of problem solving, it is necessary to analyse other components, either from a cognitive or affective perspective. Wu & Molnár (2022) discovered reasoning as a component skill of complex problem solving that are beneficial for developing strategies in a problem-solving environment. Friege & Lind (2006) emphasised the role of knowledge as a precondition to solving a certain problem, which confirmed the importance of specific knowledge and competencies in problem-solving activities. Several studies also revealed affective factors such as attitude and motivation that potentially have a direct or indirect effect on problem solving (Güven & Cabakcor, 2013; Martin, 2009). These affective factors promote individual engagement, the use of strategies, and persistence when solving problems. Both cognitive and affective factors have influenced problem solving with their respective magnitude (Shin et al., 2003). However, existing studies on influencing factors are limited to sequential investigation without considering all factors simultaneously. This hinders the possibility to explain the interconnectivity of each factor in problem solving. Indeed, a comprehensive analysis that includes multiple predictors will provide better explanations and information regarding the dynamics of problem-solving.

To sum up, the present study fills the gaps from previous research by measuring students' problem-solving ability along with their cognitive and affective factors in the Indonesian context. This is followed by an investigation of the role of cognitive and cognitive factors,

gender, and socioeconomic status in problem-solving concurrently, exploring their influence on predicting problem-solving in detail.

### **1.3 The structure of the dissertation**

To present the core information and the result of the study, we organised the dissertation into five main chapters. Chapter 1 as an introduction part, explains the rationale and context of the study. This chapter shows the research gaps and explains the importance of the study which addresses issues stated in the research goals. Chapter 2 provides the theoretical construction of the study. It explains the terminology of the variables of each measure (problem solving and cognitive and affective factors) and how those variables are connected to problem solving. This chapter also presents the evaluation of problem-solving in the Indonesian context that forms the basis of the study. Chapter 3 explains the methodology and general properties of the study. Since the study is focused on problem-solving assessment, the adaptation and development of instruments for problem-solving measurement (domain-general and -specific), cognitive (inductive reasoning and scientific competency), and affective (attitude towards science and science motivation) factors were explained along with the background information of the students. The technique of instrument validity, evaluation procedure, sample, data collection, and analysis are described thoroughly in this chapter. Chapter 4 shows the research aims, questions, and hypothesis of the study. Chapter 5 presents the empirical findings of the study. This chapter mainly addresses instrument validation results, student ability profile, and the relationship between problem solving and cognitive and affective factors. This chapter outlines seven parts, including the assessment of domain-general problem-solving, domain-specific problem-solving, inductive reasoning, scientific competency, attitude towards science, science motivation, and the relationship between cognitive and affective factors, gender, and socioeconomic status on problem solving. Chapter 6 summarises the main findings of the study and provides a theoretical explanation of the interpretation of the results. It presents the implications and limitations of the study, as well as future recommendations for further research on problem solving. I have used my previous publication for some chapters of the dissertation. The publication related to the systematic review of problem-solving assessment in Indonesia (Wicaksono & Korom, 2022b) is presented in Chapter 2. The publications related to empirical evidence of instruments' validation (Wicaksono & Korom, 2022a, 2023a, 2021) are presented in Chapters 4 and 5. Furthermore, publications explaining the relationship between variables (Wicaksono & Korom, 2023b, 2022c) are presented in Chapters 4 and 5.

## **CHAPTER 2 LITERATURE REVIEW**

This chapter provides the theoretical foundation of the study, explaining the basic understanding of each variable and its connection based on a systematic review supported by previous studies. I include several publications to explain some parts of this chapter, specifically in the theoretical aspects of problem solving (Wicaksono & Korom, 2022b) and other variables (Wicaksono & Korom, 2023a).

### **2.1 The concept of problem solving**

#### **2.1.1 Definition of a problem and solving a problem**

To get a deep understanding of problem solving, many philosophers and psychologists think about the root definition of problems and how to solve them. Based on the Cambridge dictionary, we know the word ‘problem’ as a problem or situation that is regarded unwelcome or harmful and needs to be dealt with and overcome. It can also be defined as a situation, person, or thing that needs attention and needs to be dealt with or solved. All of them refer to something that needs to be solved. The term ‘problem’ is conjugate with the word ‘solve’ the problem that mathematicians refer to as problem solving (Schoenfeld, 1987). Others argue that the problem is related to a condition, as Glaser et al. (2009) mentioned that the problem was a situation in which we were trying to reach some goals. The solution can be a goal for problems, and every problem solver is seeking it. In addition, it is important to emphasize that the problem has more conceptual depth than just a question. There should be a clash between belief and claims when thought and fact or thought and thought no longer agree (Carlson & Bloom, 2005). Problems requiring creative solutions arise not only in the process of scientific discovery but also brain teasers that may seem trivial at first but quickly make most people frustrated.

There are two classes of problems: ill-defined problems, which are characterized by the lack of solution path (Jonassen, 1997). There is no exact solution, so there are multiple solutions to such a problem, which makes solving the problem more challenging (Davidson et al., 2003). Multiple arguments and problem representation may be the best way to find the right solution for ill-defined problems. In contrary, well-defined problems, where the objectives, the way to solve the problem and the obstacles to achieving the solution are well-known, based on the available knowledge and information. Regardless of the nature of the problem, problem solving involves a cognitive process to transform a given situation where no method or solution is available (Mayer & Wittrock, 2006). In complement to this description, in the PISA assessment framework, problem solving is described as an individual’s capacity to engage in cognitive

processing to understand and resolve problem situations (OECD, 2014). It focuses on the aspects of the problems and the quality of the solution.

The process of solving a problem becomes critical, whether it is carried out through a simple, routine, familiar procedure, or requires complex cognitive operations. Many neurologists even propose a speculative model of general problem-solving brain function that arises from a series of subnetwork and systems-level interactions that together orchestrate multifaceted cognitive procedures (Bartley et al., 2018). Indeed, problem solving can be defined based on its cognitive processes. Pólya, (1945) explained that problem-solving involves some activities including understanding the problem, developing a plan, carrying out the plan, and looking back. Furthermore, solving the problem is accounted for as variant formulations of seeking the truth and building a foundation of knowledge. It is a principal unit of achievement (Nickles, 1988). Davidson et al. (2003) explained that the stages of problem solving involve the ability to recognise or identify the problem, define and present the problem mentally, develop a solution strategy, organise their knowledge about the problem, allocate mental and physical resources for solving the problem, monitor his or her progress toward the goal, and evaluate the solution for accuracy. They said that these stages are not always processed sequentially throughout the solution. Successful problem solvers are quite flexible in that sometimes solving one problem gives rise to another problem, which needs to be solved through the problem-solving cycle. Problem-solving activities work as a linear progression from one stage to the next level, where problem solvers impose an enquiry process to discover the truth, find the formulation of good problems, look for a solution for it, and test those solutions.

### **2.1.2 Domain-general and domain-specific problem-solving**

Domain-general problem-solving encompasses the varied skillset that is necessary for a person to adapt to the cross-cultural problem environment typical of contemporary societies. It touches on several cognitive and metacognitive skills such as information processing, representation, and evaluation of knowledge, reasoning, self-regulation, meta-strategic thinking, proactive planning, and decision-making (Greiff et al., 2014). The predictive value of domain-general problem-solving in educational and occupational contexts reflects the relevance of these skills in daily life in the 21<sup>st</sup> century. For example, in the secretary job, organising and scheduling a business meeting relies on domain-specific solutions, such as the standard procedure for planning and scheduling meetings and events. To do this task, a combination of factual knowledge and domain-specific problem-solving strategies will usually suffice. However, in some cases, a new situation might arise or might unexpectedly change in such a

way that the current solution is not sufficient to solve the problems. This is when domain-general problem-solving comes into play. It has the purpose of adequately communicating new situations, exploring alternative solutions, and making quick and correct decisions and adjustments (Greiff & Neubert, 2014).

The problem-solving complexity appears not only for its characteristics but also its application in a certain condition or problem situation. Despite being manifested in intelligence and general domain ability, some studies put problem-solving in a special, domain-specific, based on its context. This term appeared when Pólya (1945) used problem solving in mathematics education. He uses problem solving in starting mathematics learning and later describes problem solving in several steps, starting with understanding the problem, devising a plan, carrying out the plan, and evaluating the problem solution. Then, many educational practitioners and researchers implement problem solving in a specific subject with various learning situations (Mukhopadhyay, 2013; Yu et al., 2010). Indeed, the acquisition of problem solving for students is mostly provided by teachers through the teaching-learning process. They mainly deliver problem-solving in domain specificity based on the learning topic. The main feature of domain-specific problem-solving lies in the position of knowledge for problem solving construction. Hence, back to the core of the main function of problem solving in giving a solution to a problem, the knowledge and strong information related to the problem are needed. During problem solving activity, the solution idea mostly arises depending on the information processed (Walker et al., 2016). Wolff (2017) expresses the strong effect of knowledge-based knowledge in solving problems. This knowledge is not only specified for the content information, but also acts in organising and representing information to facilitate an efficient problem-solving process. The other study argued that problem solving is one of the human competencies that are considered domain-specific, it has a relatively narrow domain (Sternberg, 2018). As a competency, most people had mastered it in one specific domain and were less so in others. Furthermore, some studies focused on joint action in solving a problem, and the fact that their specific condition with the problem, suggesting the importance of a situated condition in problem solving. This makes an argument that if someone is going to solve one problem, they cannot guarantee that they can solve another problem in a different situation. Thus, knowledge-based problem solving can be broken down into two categories based on the information in the subject-related problem and how they use or process it. The main core of this logic is that differences between knowledge attributes (facts, theories, principles, definitions, and strategies) play a different role in the way of solving the problem starting from exploring

the context, discovering information, building a hypothesis, and confirming or verifying the solution (Csapó & Funke, 2017).

### **2.1.3 The assessment of domain-general and domain-specific problem-solving**

#### ***2.1.3.1 Assessment of complex problem-solving***

Complex problem-solving (CPS) is one of the domain-general that is characterised by the successful interaction with a dynamic task environment and gained integrated information in that process (Wüstenberg et al., 2012). It is also described as a successful interaction with non-routine and dynamic changing of environment. It represents a variety of situations that occurred in daily life (Rudolph et al., 2017). CPS emphasises a complex cognitive system, such as plan the action, develop strategy, acquire knowledge, and perform the evaluation that led to specific goals (Funke, 2010). Basic knowledge is needed to identify the most relevant structure of the problem and helps in covering possible states of the problems, as well as the problem structures and schemas. The benefit of knowledge is related to fast prediction and problem analysis, which allows the problem solver to accept, reject, or modify previous assumptions (Greiff et al., 2015). However, the main point in solving complex problems is that not all the information is necessary to solve the problem. It relates to a process in generating information processes with adequate strategies and procedural abilities to control the given system (Wüstenberg et al., 2012).

As a part of domain-general problem-solving, CPS is independent from domain-specific knowledge. Knowledge is important, but in the CPS task, a lot of information is not needed because it will result in decreasing the process of controlling the system and integrating knowledge. The requirement of a problem solver in a complex problem-solving scope includes (1) the complexity related to reducing information, (2) transparency in generating information, (3) interconnectedness in building the problem model, (4) dynamics in forecasting and controlling future development, and (5) polytely. CPS is comprised of two phases, knowledge acquisition and knowledge application (Dindar, 2018; Funke, 2010; Greiff et al., 2014). In the knowledge acquisition phase, the problem solver identifies the dynamic and variables provided by the system and tries to develop a representation of the stated problem. They need to explore and understand the complex system. In the knowledge application phase, the test takers transform the complex system into a specific state and control it by updating their knowledge.

The assessment of CPS varies between studies. In some cases, their results reported many variations and differences. Based on a meta-study in the assessment of CPS, the test was coded

into three different measurements involving classical CPS measurement, single complex system (SCS), and multiple complex systems (MCS) (Stadler et al., 2015). A classical measurement of problem solving, for example, microworlds, emulated real-world problems (Greiff et al., 2015). At the very beginning, microworlds present very broadly and multiple problem situations that are intact to society, such as governing a small town. Then it was replaced by specific problem situations in the new context “microworlds taylorshop” that represents problems in the retail business (see Figure 1). All of the variables included in the Taylorshop are very similar to real-world problems, even if the test taker manipulates those variables, it can give result in the microworld that is similar to the real condition (Greiff et al., 2015). Hence, the classical problem-solving test may be simple, but it is necessary to give a simulation of realistic problem situations. There are some limitations to the use of classical measurement in assessing problem solving. Researchers developing classical CPS measurement did not make systematic use of a common theoretical framework to produce comparable problem-solving tasks, but likely focused on real-world resemblance (Funke, 2001). Another weakness in classical measurement is that one task comprises several interrelated items. The solution of the problem is influenced by many other variables and previous actions; thus, the items are hardly independent of each other.

Month 2 of 12

Variable	Value	Planning	Variable	Value	Planning
Account Balance	168075	↑ ⓘ	Company Value	247077	↓ ⓘ
Shirts Sold	347	↓ ⓘ	Customer Interest	703	↓ ⓘ
Raw Material Price	8.27	↑ ⓘ	Raw Material in Stock	0	↓ ⓘ
Shirts in Stock	0	↓ ⓘ	Raw Material Ordered	250	<input type="text"/> + - ⓘ
Workers 50	8	<input type="text"/> + - ⓘ	Machines 50	10	<input type="text"/> + - ⓘ
Workers 100	0	<input type="text"/> + - ⓘ	Machines 100	0	<input type="text"/> + - ⓘ
Wage	1080	<input type="text"/> + - ⓘ	Maintenance	1200	<input type="text"/> + - ⓘ
Shirt Price	52	<input type="text"/> + - ⓘ	Social Events	50	<input type="text"/> + - ⓘ
Retail Stores	1	<input type="text"/> + - ⓘ	Advertising	2800	<input type="text"/> + - ⓘ
Worker Satisfaction %	57.7	↓ ⓘ	Business Location	City	City ⓘ
Loss of Production %	26	ⓘ	Machine Damage %	11.2	↓ ⓘ

Commit

Figure 1. The example of taylorshop task (Greiff, Stadler, et al., 2015, p. 103)

In addition to classical measurement, problem-solving tests have also been developed in the form of a single complex system (SCS). One of the popular SCS problem-solving tests is MultiFlux, which was firstly designed by Kroner et al. (2005) based on the linear structural equation system (Figure 2). This test is considered as a one-item test because the scenarios are generally constructed by a specific system configuration. Indeed, every indicator that assesses

every domain in this test during system exploration is related to the same structure (Wüstenberg et al., 2012). Even when the test taker performed the different tasks in the test series with different goals, it still depends on the same structure of the system. In this programme, participants will explore some tasks that have an additional effect on the generation of knowledge. This test assesses problem solving while minimising the undesirable effect of domain-specific prior knowledge that was set in embedding an overarching cover story centred around exploring and controlling fictitious machine (Christ et al., 2020). The MultiFlux has four principles in their system, including rule identification strategy, causal knowledge, rule knowledge, and rule application (Wüstenberg et al., 2012).

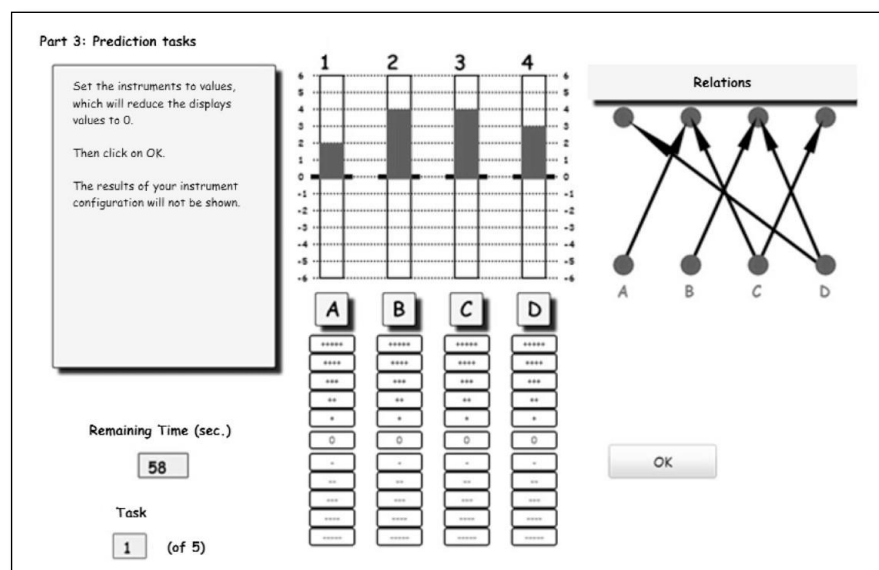


Figure 2. The example of multiflux task (Christ et al., 2020, p. 4)

The third type of CPS is the multiple complex system (MCS), which has some differences compared to classical and SCS in terms of the variables. MCS focuses on psychometric properties and the core features of CPS (Greiff et al., 2015). In contrast to classical and SCS problem solving, MSCs use multiple and independent items to assess problem-solving ability (Wüstenberg et al., 2012). One of the MCS problem-solving tests that is widely used in the assessment process is MicroDYN. This test is a computer-based assessment with multiple independent elements and multiple control roles in each element (Greiff, 2012). This test consists of 8–10 complex items with 3 different input variables (denoted as A, B, C) and three output variables (denoted X, Y, and Z) (Rudolph et al., 2017). The input variables influence the output variable so that only the input variables can be actively manipulated by the problem solver. Thus, there are two types of interconnections between the variables in which input variables influence the output variable or output variables influence each other or themselves

(eigendynamic). As is seen in Figure 3 for the example, the input variable 'A' can influence the output variable 'X', while the variable 'B' influences the variable Y or Z. Even the same output variable, such as the variable 'Y', can influence the variable 'Z'. Then, variable C influences variable Z, and so on. This complex interrelation between variables is worked on based on the specific equation in which all possible relations are equal to the number of the output variables.

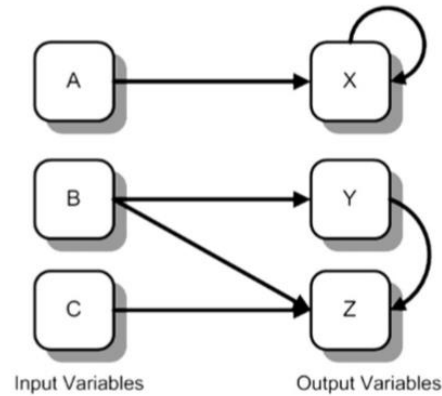


Figure 3. The scheme of variable involved in the MicroDYN test (Wüstenberg et al., 2012, p. 4)

MicroDYN, as one of the online-based CPS tests, allows test takers to use multiple independent items and multiple control rounds within their items (Schweizer et al., 2013). Since test takers must find the relationship between the input and output variables, the MicroDYN framework was constructed based on linear structural equations (Greiff et al., 2012). Test administration was started with a brief explanation about general instruction and test trials to make test takers familiar with the handling and other components of the test. All items in the test are manifested by two different phases, knowledge acquisition and knowledge application. In the knowledge acquisition phase, test takers identify and explore the system by modifying the input variables and verify their effect on the output variable indicated by the change in their value. Here, their strategy is assessed based on their actions to understand the system structure (facet rule identification). At the same time, test takers must note the relationship between the input and output variables by drawing a line to connect both related variables. In this phase, the knowledge of the facet rules was assessed as the correct result of the causal relationship model. In the knowledge application phase, the test takers will face the same items, but they must reach a specific value in the output variable by manipulating the input variables. In this phase, test takers were given a correct causal relationship between the input and output variables before completing the test. The test takers must reach a certain value in the input variable with no more than four control rounds. The application of facet rules was evaluated in this phase. As an

example, Figure 4 illustrates the MicroDYN task. In this test, students are presented with three input variables related to cat food (brekon and mikas) and two output variables for cat behaviour (purring and movement). The aim of the task is to explore and find the relationship between input variables (cat food) and output variables (cat behaviour). In the first phase, students explore the system by manipulating or controlling the value of input variables and observing the changing in the output variables. Through these activities, they are expected to figure out the connection between variables and represent it in a model. In the second phase, students are asked to reach the target value in the output variable. Consequently, they must accurately adjust the value of each input variable.

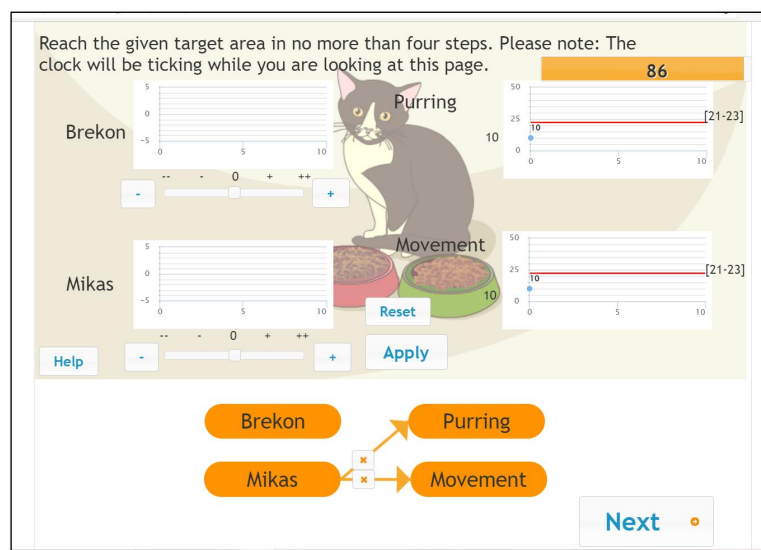


Figure 4. Screen capture of the MicroDYN task in “cat’s food and behaviour” during the application phase (Wu & Molnár, 2021, p. 1016)

### 2.1.3.2 Assessment of domain-specific problem-solving in science

The problem-solving test in specific domain was raised since the invention of mathematics problem-solving which led to the development of research in problem solving and scientific inquiries (Schoenfeld, 1987). The first problem-solving test was assigned several sets of indicators, including understanding the problems, devising a plan, carrying out the plan, and evaluating the problem solutions (Pólya, 1945). Later, the problem-solving test developed in different contexts like science with the use of those indicators and presented in science problems and phenomena. The progression of the problem-solving framework also occurs due to psychological approaches and cognitive theory. The framework was changing, and information was adapted by several researchers. As an example, Rausch & Wuttke (2016) documented the phases of a problem-solving approach. They identify four main steps including (1) identifying

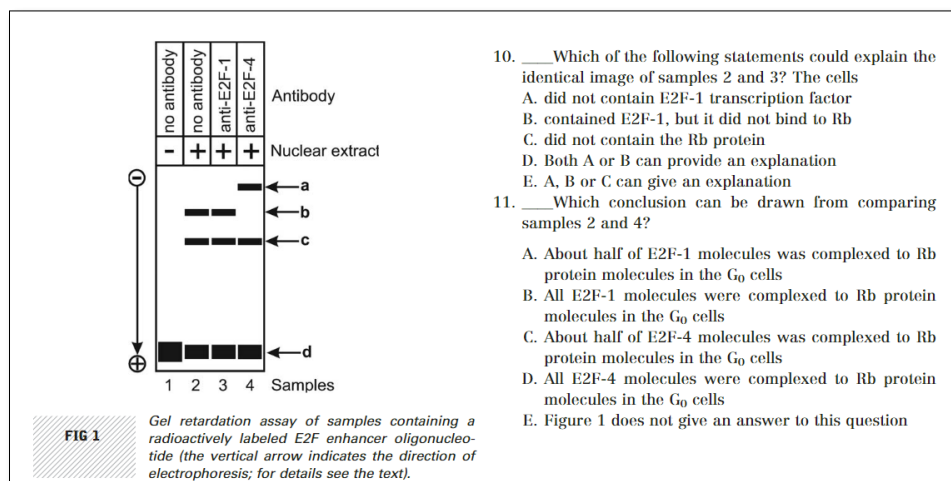
information gaps and needs for action, (2) processing information when information and knowledge related problems are stored, interpreted and used to understand the problem and make a decision (Lachman et al., 2015), (3) coming to a well-established solution. Based on available information and cognitive processing, a solution can be proposed with strong analytical calculation and (4) communicating the decision, although the problem is solved, it is necessary to communicate the result of the solution of the problem. Hence, communicating a solution in oral or written form and making anyone aware of it becomes an important facet of domain-specific problem-solving. Schoenfeld (2013) explains that if a person wants to engage in a goal-orientated activity, such as problem solving, he or she needs to make a series of activities, such as stating the goals, maintaining individual knowledge or resources of his or her disposal, developing individual beliefs and orientation, and making decisions.

The other approach came from Singer & Voica (2013), which described problem solving in four phases including decoding (transporting text into compatible data), representing (transposing problem through mental model), processing (identifying mental model and mental configuration based on the identified model and personal competencies), and implementing (applying techniques for a particular problem to generate solutions). Similarly, the other study in STEM education also categorised problem solving into four distinct variables with the main focus on the monitoring and self-regulation approach. The four variables include analysing the task, creating a plan, executing the plan, and monitoring understanding and actions (Phillips et al., 2016). The diversity of problem-solving frameworks mostly lies within the focus of assessment activities, in which some frameworks emphasise data information suitable for problem solving related to data science. On the other hand, some frameworks focused on monitoring activities and empowering the self-regulation system. Thus, all approaches to problem solving can be varied in their construct based on the context and needs, but the core concept of problem solving should be addressed by involving components of the problem and solutions.

Domain-specific problem-solving has a different position in real-life problem-solving activities. Although domain-general problem-solving plays an important role in the daily routine real-world problem-solving context as basic intelligence skills, domain-specific problem-solving is mostly used in educational training for teaching problem solving. Also, for a complex and specific problem situation, the procedure for making solution and decision is the same as the specific domain problem-solving approach. That makes sense, in the real world, a complex problem should be given by experts to solve it. In fact, they also use a general domain framework to solve the problem efficiently. In conclusion, both general and specific problem

solving are needed in solving a real-world problem when domain-general plays as basic intelligence in individual skills and specific domain support in providing a problem-solving approach with comprehensive knowledge and information.


The problem-solving measurement task can be constructed in different forms and formats. One example of a problem-solving test on a science topic (biochemistry and molecular biology) used multiple-choice completions for the assessment (Bátor & Szeberényi, 2021; Szeberényi, 2013). This test presented problems using lab experiment scenarios with additional explanations of the experiment's procedures, the data, and the results. The students must find the right answer to the questions by identifying the experimental procedures and data. Thus, this type of problem-solving test is based on knowledge and targets a specific population, such as medical and science students (see Figure 5). At the high school level, the problem-solving test was found in the form of open-ended questions and multiple choices. The study conducted by Fortus et al., (2005) presented well-defined problem-solving test in the scope of earth science (Figure 6). They focus on assessing problem-solving ability with the support of prior knowledge and representations relevant to the problem.



10. \_\_\_ Which of the following statements could explain the identical image of samples 2 and 3? The cells
- A. did not contain E2F-1 transcription factor
  - B. contained E2F-1, but it did not bind to Rb
  - C. did not contain the Rb protein
  - D. Both A or B can provide an explanation
  - E. A, B or C can give an explanation
11. \_\_\_ Which conclusion can be drawn from comparing samples 2 and 4?
- A. About half of E2F-1 molecules was complexed to Rb protein molecules in the G<sub>0</sub> cells
  - B. All E2F-1 molecules were complexed to Rb protein molecules in the G<sub>0</sub> cells
  - C. About half of E2F-4 molecules was complexed to Rb protein molecules in the G<sub>0</sub> cells
  - D. All E2F-4 molecules were complexed to Rb protein molecules in the G<sub>0</sub> cells
  - E. Figure 1 does not give an answer to this question

Figure 5. The items sample of the biology problem-solving test (Szeberényi, 2014, p. 89)

19 An igloo is a structure that is used for survival in extremely cold environments with snowstorms. The structure is typically made of blocks of ice laid one on another in order to form the shape of a dome. Describe how you would test this structure to collect data on its ability to withstand static forces and dynamic forces, and how well it prevents heat loss.



(a) Static and dynamic force test description:  
\_\_\_\_\_

(b) Prevent heat loss test description:  
\_\_\_\_\_

Figure 6. The items sample of the science problem-solving test in high school (Fortus et al., 2005, p. 878)

### 2.1.3.3 International large-scale assessment of problem-solving

In the domain-general problem-solving, one of the most well-known assessments is creative and collaborative problem-solving conducted by the Programme for International Student Assessment (PISA). They have developed assessment tools to compare the educational system worldwide (OECD, 2016). In 2012, PISA introduced a cognitive problem-solving assessment framework with a computer-based test design in which students face daily routine real-life problems such as using a new mobile phone, how to fix problems with lamps and electricity, or finding some location with a different path/route (OECD, 2014). However, they involved nonroutine tasks, but the problems are pointed to general knowledge strategy. The PISA creative problem-solving framework is presented in three distinct aspects: the nature of problem situations, the problem-solving process, and the problem context (Table 1).

The PISA creative problem-solving test consists of static and interactive tasks. The static tasks focus mainly on decision-making problem tasks with different types of static units. All units are based on computer video game mechanics (Dindar, 2018). The interactive unit in the PISA test belongs to complex problem-solving, MicroDYN, and finite-state automata (OECD, 2014). They use the 'control' and 'exploration' of an unknown system for student problem-solving tasks. Four units in the test are MicroDYN units, and six of them are finite-state automata. The measurement of the PISA creative problem-solving test is classified into seven proficiency levels (below 1, 1, 2, 3, 4, 5, 6, and 7) based on the items or units for which the test taker can solve. The lowest level, below 1, is presented by the 'traffic' task. In which the ability of the test taker problem-solving proficiency is about the ability to read the travel time of a

simple network diagram to find the shortest route between two close points on a map. All necessary information is disclosed at the beginning of the report and response options are provided. The correct solution can be found with a few simple trial-and-error iterations. The highest level is presented by the ‘robot cleaner’ task (see Figure 7). If the test taker reaches this level, he/she will acquire the highest problem-solving skills in which he/she fully describes the logic governing an unfamiliar system. After observing the behaviour of a simulated robot cleaner, the student identified and wrote down the two rules that, together, completely describe what a robot cleaner does when it meets certain kinds of obstacles.

Table 1. The description of the creative problem-solving framework in PISA 2012 (OECD, 2014, p. 126)

Main aspects	Details
The nature of problem situations	Interactive: Not all information is disclosed; some information must be uncovered by exploring the problem situation
	Static: All relevant information to solve the problems is disclosed at the outset
Problem-solving process	Exploring and understanding: The information provided with the problems
	Representing and formulating: Constructing a graphical, tabular, symbolic, or verbal representation of the problem situation and formulated hypothesis about the relevant factor and the relationship between them
	Planning and executing: Devising plan by setting goals and subgoals and executing the sequential steps identified in the plan
	Monitoring and reflecting: Monitor progress, react to feedback, and reflect on the solution, the information provided with the problem, or the strategy adopted.
Problem context	Setting: Technological devices (use technology or non-technology) embedded in problem scenario
	Focus: The subject of problems related to (personal or social)

Later, a different type of problem-solving test, collaborative problem-solving, was developed by incorporating creative problem-solving tasks with team collaboration features (OECD, 2017a; Pásztor-Kovács et al., 2023). This construct has a rationale based on workplace demand, in which problem-solving is not needed for an individual task only, but for a whole team project. The problem itself presents a common goal for individuals. In a task in which no single individual can solve the problem independently, collaborating with others becomes essential for a successful result. This collaboration combines methods of joining resources in response to the analysis of the problem space (Care et al., 2016). Even in the school environment, collaborative interaction between students in class performance will result in better achievement and their ability in solving tasks (Fawcett & Garton, 2005).

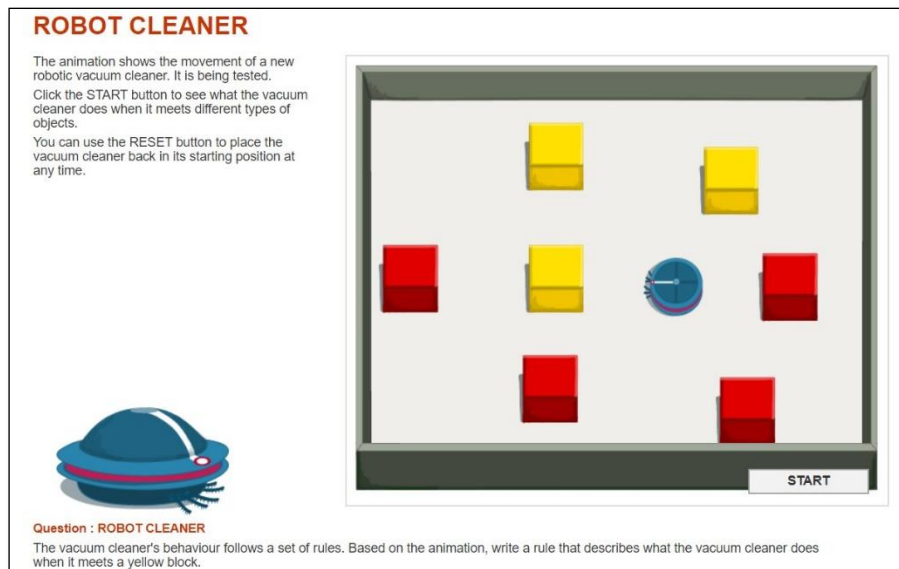


Figure 7. The example of PISA problem solving test; showing the task to identify the work pattern of robot cleaner (OECD, 2014, p. 42).

#### 2.1.4 The assessment of problem solving in Indonesia

The evaluation of problem-solving skills has been done in many countries, since PISA launched a collaborative cognitive problem-solving assessment. In Indonesia, the lack of participation in PISA or other global problem-solving assessment programmes result in limited information about students' problem-solving ability compared to other countries. Despite many studies conducted on problem-based learning to improve student problem-solving skills in Indonesia (Asyari et al., 2016; Iswandari et al., 2017), only few studies have been found on assessment area. The review of the literature on problem-solving assessment in Indonesia was done by Wicaksono & Korom (2022) to evaluate the problem-solving assessment developed in the Indonesian version. The study on problem-solving test was collected by conducting a comprehensive search in the databases DOAJ, Research Gate, ERIC, and Google Scholar published from January 2010 to December 2019. The search strategy is varied between databases, but it commonly includes keywords such as "assessment", "test", "problem-solving", "mathematics and science", "validation/validity", or "Indonesia". Systematically a search was conducted by entering the combination of keywords in the databases in English and Indonesian. All the studies collected were conducted in Indonesia and administered the Indonesian version of problem-solving tests. All published studies in journal and conference proceedings are included in the review. All studies included in the review met the following criteria: (1) the tests are standardised; (2) the sample size of the study was reported; (3) having empirical results in validity and reliability; (4) the tests are projected to 7-12 grade and university students. The

total amount of 93 studies on the Indonesian problem-solving test was found. 61 (65.59%) of them were excluded because the study only showed the test design without any report on the empirical study. In the end, a total number of 32 (34.41%) studies were included with criteria to show the results of the structure, validity, and reliability of the test. The features related to the focus of the study were coded including: (a) the topic/content of the test focus (i.e., mathematics, science, physics, chemistry, or biology), (b) the test development quality that indicated by validity and reliability value, (c) grade distribution, (d) test framework, indicate by the theoretical background used for developing test items, and (e) sample size.

Out of the 32 published studies used in the review, 1 study (3.13%) was found in 2012, followed by 2 (6.25%) studies in 2014, 1 (3.13%) in 2015, 4 (12.50%) in 2016, 7 (21.87%) in 2017, 9 (25.12%) in 2018, 5 (15.63%) in 2019, and 3 (9.37%) in 2020. The tests used a specific subject and a different problem-solving framework. Each problem-solving framework is translated into multiple choice and essay-form questions. The total of items developed in the studies ranged from 5 to 103 items, where all tests are in paper form. The details of the problem-solving test founded in Indonesia are described in Table 2.

In developing a problem-solving test, each study used different frameworks and indicators. Most of the study used Polya's problem-solving framework in which the items represent skills to understanding the problem, devising a plan, carrying out the plan, and evaluates the problem solution. It recorded that 11 of 32 studies used Polya's problem-solving framework and were translated into one or more different questions. The other 3 studies used Doctor and Heller's problem-solving framework that has a specific term in physics or science context. 3 studies used the OECD cognitive problem-solving framework that was modified to specific mathematics and chemistry questions. Furthermore, nine studies used different references to construct test items to solve problems. There is one study that did not explicitly mention their main references but only gave an explanation that they used problem-solving indicators in choosing a strategy to solve mathematical problem-solving. The last five studies did not explain their construct framework explicitly and only mentioned 'solving the problems'. The detailed problem-solving framework used in the literature study is shown in Table 3.

Table 2. The problem-solving test developed in Indonesia

No	Author (s)	Year	Subject	Test items		Number of participants	Reliabilit y	Type of validity
				$\Sigma$	type			
1	Sinaga	2016	mathematics	5	essay	90	.650	Content & Construct
2	Pardimin et al.	2017	mathematics	5	essay	90	.803	Content
3	Anggraeni et al.	2018	mathematics	15	essay	40	.732	Content
4	Wahyuningrum	2014	mathematics	5	essay	122	.595*	Content & Construct
5	Zulkarnain et al.	2018	mathematics	5	essay	30	.690	Content & Construct
6	Ariawan	2016	mathematics	6	essay	78	.640	Content & Construct
7	Bidasari	2017	mathematics	12	essay	36	.871	Content & Construct
8	Novita	2012	mathematics	15	essay	25	.737	Content
9	Februarini et al.	2017	mathematics	6	essay	38	.800	Content & Construct
10	Putra	2017	mathematics	5	essay	10	.580*	Content & Construct
11	Subekti et al.	2014	mathematics	5	essay	29	.761	Content & Construct
12	Hendriyana et al.	2018	mathematics	5	essay	66	.884	Content & Construct
13	Peranginangin et al.	2019	mathematics	4	essay	32	.751	Content
14	Irawati et al.	2018	mathematics	4	essay	5	.780	Content & Construct
15	Alfika et al.	2018	science	5	essay	10	.697	Content
16	Hidayat et al.	2017	science	15	essay	35	.880	Content & Construct
17	Lutfi et al.	2019	science	15	essay	30	.835	Content & Construct
18	Pratiwi et al.	2015	physics	15	essay	35	.917	Content
19	Nadapdap et al.	2017	physics	52	multiple choices	281	.800	Content & Construct
20	Syifauliyah	2019	physics	10	essay	30	.880	Content
21	Lestari et al.	2019	physics	5	essay	40	.839	Content & Construct
22	Sinensis et al.	2019	physics	20	essay	52	.759	Content & Construct
23	Kurniawan et al.	2018	physics	8	essay	60	.670	Content & Construct
24	Yulianti et al.	2018	physics	10	multiple choices	34	.780	Content & Construct
25	Sutiadi et al.	2016	physics	24	essay	37	.431*	Content & Construct
26	Rifa'i et al.	2018	physics	5	essay	101	.621	Content
27	Savitri et al.	2020	physics	5	essay	25	.770	Content & Construct
28	Wardhani et al.	2020	physics	5	essay	25	.800	Content & Construct
29	Faturrahman et al.	2016	chemistry	10	essay	37	.730	Content
30	Harta	2017	chemistry	5	essay	127	.930	Content & Construct
31	Hidayat et al.	2018	chemistry	4	essay	42	.710	Content
32	Ridhwan et al.	2020	biology	10	essay	136	.876	Content & Construct

Note: The reliability value shown in the table uses Cronbach-alpha reliability. \*the past reliability cannot be considered as high (e.g. above .80) or at least marginally acceptable (e.g. above .60) (Gliner et al., 2016).

Table 3. The Problem-solving frameworks used in the literature study

Authors	Problem-solving aspects
Subekti et al. (2014); Alfika et al. (2018); Pardimin et al. (2017); Wahyuningrum (2014); Pratiwi et al. (2015); Nadapdap et al. (2017); Lestari et al. (2019); Anggriani et al. (2018); Hendriyana et al. (2018); Irawati et al. (2018); Wardhani et al. (2020);	(a) understanding the problem, (b) devising a plan, (c) carrying out the plan, and (d) evaluating the solution of the problem (Pólya, 1945).
Hidayat et al. (2017); Syifaulyah (2019); Savitri et al. (2020)	(a) Visualising the problem, (b) using physics approach, (c) making a specific application in physics, (d) using a mathematical procedure, and (e) defining logic conclusion (Docktor & Heller, 2009)
Bidasari (2017); Novita (2012); Hidayat et al. (2018)	(a) exploring and understanding, (b) representing and formulating, (c) planning and executing, (d) monitoring and reflecting (OECD, 2014)
Putra (2017)	(a) Formulating the problem, (b) defining semistructural information, and (c) achieving the solution (Sumarmo, 2015)
Kurniawan et al. (2018)	(a) combining imaginative skills, (b) developing a model, (c) conducting an investigation, (d) analysing data and making conclusions (Butterworth & Thwaites, 2013)
Sinensis et al. (2019)	(a) Problem schema, (b) analogy, (c) causal, (d) argumentation (Jonassen, 2010)
Luthfi et al. (2019); Sutiadi et al. (2016); Peranginangin et al. (2019)	(a) identifying the problem, (b) determining goals, (c) making strategy, (d) exploring strategy, (e) making action (Brookhart & Nitko, 2014)
Faturrahman et al. (2016); Ridhwan et al. (2020)	(a) identify the problem, (b) define and represent the problem, (c) explore possible strategies, (d) act on the strategies, and (e) look back and evaluate the effects of your activities (Bransford et al., 1986)
Yulianti et al. (2018)	(a) Scientific approach, (b) structured and unstructured manner, (c) memory-based approach, and (d) no clear approach (Walsh, et al., 2007)
Februarini et al. (2017)	Choosing a strategy to solve the problems.
Ariawan (2016); Zulkarnain et al. (2018); Sinaga (2016); Harta (2017); Rifa'i et al. (2018)	Solving the problem

The problem-solving test developed in the Indonesian context strictly follows the national curriculum regulation and is administered as core competency in the curriculum. The tests

addressed specific topics and targeted different grades. Of the 32 studies collected in this review, they specifically addressed a certain grade. Starting from middle school from grade 7 to grade 9, 10 studies are founded for the mathematics problem-solving test and 4 studies for the science problem-solving test. In high school, problem-solving tests are available for mathematics (4 studies), physics (4 studies), chemistry (3 studies), and biology (1 study). Then for the undergraduate level, the problem-solving test is only found in physics (5 studies) and mathematics (1 study) courses. The topic used in the problem-solving test varied across the grades. The subtopic and subject competency are different from grades 7 to 12. For the undergraduate level, there is no restriction because every higher educational institution can construct and modify its curriculum. There is a general standard for higher education in Indonesia provided by the national education department, but every institution or university has the authority to make and implement its educational system. The detailed topics of the problem-solving items are given in Table 4.

Table 4. Topic distribution of problem-solving assessment tools in Indonesia

Grade	Mathematics	Science		
7	Set and number, linear function, geometry.	Environmental issue		
8	Statistic and probability, linear function, geometry	Force and motion, wave and sound		
9	Set and number, geometry, algebra, statistics	-		
	Mathematics	Physics	Chemistry	Biology
10	Trigonometry	Work and energy	Periodic table and chemical reaction	Environment and pollution
11	Function	Wave and sound, elasticity, optical instruments, static fluid, temperature, and heat	Stoichiometry, Acids and bases	-
12	Set and number	-	-	-
Undergraduate	Calculus	Thermodynamics; Dynamic electricity; momentum and impulse	-	-

The review also investigates the empirical validity and reliability test used in the study. The developed tests used Cronbach alpha reliability to check the item consistency, and it was done by SPSS software. From the reliability result, only three tests have a low-reliability score ( $r <$

.60). 19 tests have an acceptable reliability score ( $r$ -value between .60 and .80) and 10 studies reported a high reliability score ( $r > .80$ ). In term of validity, 10 studies only reported the result of content validity and 22 studies reported both contents and construct validity. The content validity conducted by the studies was focused on the topic consistency in which the item's writing is correct based on the knowledge background and the language composition. The number of experts involved in content validation ranged from 2 to 7 people. Many of them are university and high school teachers who have experience teaching related subjects. All studies were taken into account and made judgments based on expert evaluations. Therefore, some revisions were made until all items were considered valid by experts.

The construct validity was done by 22 studies that showed valid evidence of the developed items. Most studies, 19 of them, used Pearson's correlation analysis to measure the item's validity. They showed validity with good results, high positive  $r$  value and significant statistics. The  $r$  value is varied between studies, the lowest value is about  $r = .38$  the highest is  $r = .88$ . All tests in the studies are considered acceptable and valid. Furthermore, 3 studies used Rasch analysis to determine the validity of the item. They measure infit MNSQ to check the fitting items with the model. The tests showed that the infit MNSQ index ranges from 0.99 to 1.03 with an acceptable range of 0.7 to 1.3 (Griffin, 1999). That result means that all items measure problem-solving skills correctly. All problem-solving tests developed in Indonesia are within the scope of certain content knowledge and subject (mathematics, physics, chemistry, biology, and science for junior high school). The knowledge acquired from experience to the current situation will help the individual generate a solution to a problem. Depending on the nature of a problem, different knowledge plays a different role in the problem-solving process because problems occur with different conditions (Liao, 2002). Due to its preference for a specific context, this assessment contributes greatly to promoting problem-solving skills in educational practise. The tests are designed as a part of the teaching and learning process and the context of the test is matched with the core competency of the curriculum that the students are learning about. This test also serves as an evaluation of student competency both in problem-solving skills and in subject course.

Problem-solving tests are available for high school and university students. All tests were designed for specific domains and most of them used the Pólya's problem-solving framework. In this case, Pólya's problem solving that originally developed in the 1960s is constructed for mathematics education. He discovered that the knowledge in mathematics is obtainable by the use of suitable problem situations and rediscovery is a useful tool for active learning (Voskoglou, 2011). By time, this framework is developing and being used not only for

mathematics problem-solving but also for science context. Moreover, the other important issues related to the problem-solving framework is the use of the general domain problem-solving framework in the construction of specific domain problem-solving test. Some studies used the PISA problem-solving framework for mathematics and chemistry problem solving tests. In the PISA test of problem solving, the construction of their frameworks is implemented in general tasks and not connected to any curriculum subject (i.e., traffic, climate control, and robot cleaner). The idea of using domain-general framework in specific-domain problem-solving raises the universality of knowledge and principles in assessment studies.

The other studies used different problem-solving frameworks but share similar principles. For example, the first step in the problem-solving process in every study is related to understanding the problem. It is a common task when someone tries to solve a problem. They need to know what the problem is, what variables are related to the problem, and how they understand the problem. Knowing the problem will lead to a clear-thinking path and direction for stepping in the solution. Second, it is making a strategy of a plan for solving the strategy. All the tests implement this aspect. Some of them put different additional skills like determining goals and gathering information before making a strategy. The next steps are to act on the strategy or execute the plans they have made for the solution of the problem, and the last is to make an evaluation or reflexion based on the impact of the solution. However, one literature background used by the test developer is different from the others. Despite determining the process of making a solution, the problem-solving framework constructed by (Jonassen, 2010) is talking about analogy, causal relationship, and argumentation that are closer to a mental process. Those skills are important as individual thinking skills in finding a solution to a given problem.

Although the test administered the topic in curriculum, in each grade, the tests are covered by the developed tests. For a subject like mathematics, it can be found that the problem-solving test developed in every grade ranged from middle school to undergraduate level. Then for physics, there is only need for the test for twelve-grade students, and in science, the test needs to be for ninth grade. However, only a few problem-solving tests were developed in chemistry (grade 10<sup>th</sup> and 11<sup>st</sup>) and biology subjects (grade 10<sup>th</sup> in environmental and pollution topic). It is quite interesting to be talking about problem solving in chemistry and biology because those subjects have many connections to real 'problematic' situations. Many aspects in chemistry and biology represent problem situations; for example, chemical contamination, conservation, and pandemic disease are the real problems that need to be solved in the recent day. Then,

introducing students with these chemistry and biology-related problems through learning and assessment will prepare them to be future problem solvers, despite other subjects.

In the empirical test analysis, some items show acceptable Cronbach-alpha reliability. Only three developed tests have a low  $r$  value. In case of the test that has low reliability, it is better if they do some modification or revision in those items, then rerun the analysis. Moreover, for essay type test, the scoring is dependent on the strength of rubric and the raters, thus it is suggested to use additional reliability test to check the consistency of the items based on rating system such as interrater reliability (Gliner et al., 2016). In the test validity, almost half of the tests only conduct content validity which documented the relation of test specification with their content (Downing & Haladyna, 1997). This makes the test only checked by the expert and validated by the personal judgment. They checked the domain used to constitute the construct. It did not show the empirical result that reflects the position of the items based on the perspective of the test taker. Empirical investigation is critical for high-stakes examination, such as problem-solving skills, to make sure that the items correctly measure the student's skills. For the other tests, which used both content and construct validity, it showed a more solid result because they used empirical data that refer to what extent the items measure the construct (Westen & Rosenthal, 2003).

The studies used different sample numbers, but they were all relatively small (21 studies used less than 50 samples, 6 studies with 50-100 samples and 5 studies with more than 100 samples). Some references said that the more samples used in the validity measure, the better the quality of the research result. Schumacker & Lomax (2014) also mentioned that the sample numbers for conducting validity are disparately ranging from 150 to 1000 samples based on the estimated parameter method and data normality. However, there is no exact number of samples or participants that must be included in the item's validation, yet it depends on many factors. Indeed, it would be better if the sample number is relatively high and represents the exact population. Overall, the problem-solving test developed in Indonesia varied in level and topic. However, some limitations were still found in the case of psychometric properties and sample size. Further recommendation is suggested to comprehend the test analysis using factor analysis to check the sustainability of the test's theoretical framework and students' response, as well as Rasch analysis to check the fit of each item in the test. The increased number of participants to conduct pilot testing and psychometric analysis of the test was suggested to obtain a strong statistical result. Indeed, it is also possible to develop or adapt new problem-solving tests in Indonesian context, since there are some topics that have not been covered by the existing test, such as biology and science problem-solving test or even domain-general problem-solving test.

### 2.1.5 The role of gender and socioeconomic status in problem solving

The study of gender differences in problem solving showed diversity in the result and interpretation based on its context. In mathematics problem solving, there is a small difference between gender performance that favours boys over girls (Brunner et al., 2008). In physics subject, male students showed better problem-solving strategies compared to female students (Gok, 2014). Wüstenberg et al. (2014) conducted the cross-national analysis for gender differences in complex problem-solving, founding that male students from two different countries have no significant difference in complex problem-solving performance ( $p > .05$ ,  $d = .13$ ), but the significant result was found in comparison between males and females from the same country ( $p < .001$ ,  $d = .20$ ) and males and females from different country ( $p < .001$ ,  $d = .54$ ). Thus strengthening the dominance of male students in complex problem-solving performance. The PISA survey on creative problem-solving also found that male students outperformed female students in majority of participating countries (OECD, 2014). Additionally, the variation observed in boys is also larger than the girls, meaning that the ability level of male students is more diverse from low to high category. Borgonovi & Greiff (2020) explained these results in connection with societal influences, noting that society's different trajectories for males and females could shape gender disparities. Wittmann & Hattrup (2004) explained that male relies on more efficient strategies for completing problem-solving tasks, which allows them to detect variables and system structure clearly. The different result was found that there were no significant differences between male and female in domain-specific problem-solving (Gallagher et al., 2000; Gunawan et al., 2020; Guven & Cabakcor, 2013; Harskamp et al., 2008; Ramírez-Uclés & Ramírez-Uclés, 2020). Dindar (2018) also found no significant effect of gender on problem-solving ability in high school students in Turkey. A different result was found in science problem solving which showed that female students have higher activities in problem-solving process compared to male students, especially when it comes to text or diagram comparison (Huang & Chen, 2016).

The other factors, such as socioeconomic status potentially affect students' differences in problem-solving performance. This claim is supported by studies that explain how cultural and economic background affect people behaviour (Salazar & Salas, 2013) as well as the organisation skills and initiative planning process which is important for problem-solving process. OECD (2014) explained that socioeconomic status reflects circumstances where fairness and privilege barriers make it easier for some students to get better achievement. Furthermore, these individual circumstances may contribute to shaping students' motivation, aspirations, and further influence their cognitive skills and abilities. The empirical result

exploring the effect of socioeconomic status (SES) in problem solving was reported by Chiu (2022) who found a significant correlation between SES and collaborative problem solving. Martin et al. (2012) also found that students from low socioeconomic background performed lower in problem solving rather than students with strong socioeconomic status. PISA survey also found that 10.6% variation in creative problem-solving results was explained by socioeconomic status. Furthermore, the survey results showed that the performance gap related to parent's occupational status had a high standard deviation (48%), where the students from higher occupational families tend to perform better in problem solving than those from lower occupational families. Therefore, the effect of socioeconomic status was not directly associated with students' performance, but rather with unequal access to better schools and teacher quality (OECD, 2014).

## **2.2 The cognitive factors affecting problem solving**

Problem solving involves individual cognitive processes to understand the problem and provide a solution. These processes manifest in many factors that support problem-solving in respective parts. Stadler et al. (2015) confirmed that problem solving requires the involvement of personal intelligence capabilities such as reasoning. Furthermore, Park & Lee (2004) explained that problem solving requires higher thinking skills and mastery of context, which can help problem representation. Shin et al. (2003) also found that domain knowledge has a significant contribution in predicting well-defined problem-solving, explaining that individual with higher domain knowledge will be more successful in solving problems. Thus, cognitive factors are essential to explaining the dynamic of problem-solving processes. This study identifies inductive reasoning and scientific competency as potential predictors for problem solving and evaluates their contribution to predict problem-solving performance.

### **2.2.1 Inductive reasoning**

Reasoning is one of mental representation served as a basic ingredient of cognitive system mainly for enhancing thinking process and decision making (Mercier & Sperber, 2011). The mechanism of reasoning is intuitive belief from the source of information that connects the premises and conclusions. The concept of reasoning was explained Evans (2006) with heuristic-analysis theory composed of two cognitive processes: heuristic which generates specific representations of problems and analytic which refers to judgement from the representations. When people get exposed to situations, they generate mental models and hypotheses to refer to

the tendency to contextualise all problems based on their knowledge references. This stage will perform mental stimulation to forecast or predict future activities. Mental stimulation triggers the analytic system to evaluate the current hypothesis. This model (see Figure 8) conceals a degree of conditional belief and is rooted in a causal relationship.

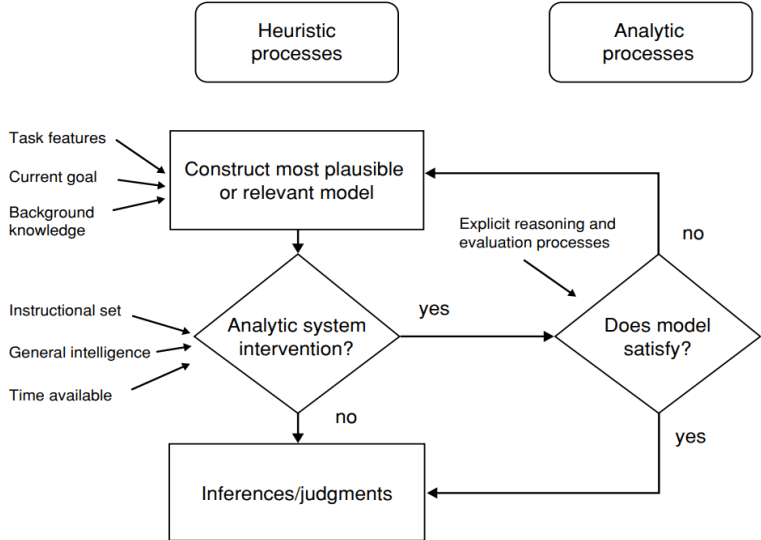


Figure 8. The heuristic-analytic theory of reasoning (Evans, 2006, p. 380)

This theory specifically discusses the hypothetic thinking which considers the analytic process as only situation involved in controlling behaviour (Evans et al., 2003). The heuristic process can generate model or response which is evaluated by analytic process to produce inferences. Indeed, the strong feature of this system is the emphasis on the ability of the analytic system to inhibit heuristic responses.

The principle model of reasoning was formed into two terms, *deduction*, as an inference from universal premises into a conclusion, and *induction*, as an inference from one particular assertion to a universal term (Johnson-Laird & Khemlani, 2013). As we focused on the inductive side, Hayes et al. (2010) explained inductive reasoning as the ability to make predictions about a new situation based on previous knowledge. The focus of inductive reasoning is generalising the property of premises into a general conclusion. Furthermore, McAbee et al. (2017) described inductive reasoning ability as the formulation of theories by drawing general inferences from empirical data. Inductive reasoning is related to cognitive processes such as categorisation, inference, probability judgment, and decision making. That cognitive processes involve the property of generalisation which is driven by the relation between premise and conclusion.

The theory for inductive reasoning process is focused on the transfer of knowledge (Csapó, 1997). The induction process improves knowledge acquisition and its application in a new situation. The knowledge transfer process is recognised as a general component of intelligence. Most inductive processes are focused on discovering the phenomena and logically finding the regularities. This process requires certain strategies such as engaging activity and experience with particular cases, formulating the logic, conjecture proof, and evaluating and justifying the cases (Christou & Papageorgiou, 2007). An additional explanation supporting this claim is that mental activities that present inductive reasoning detect generalisation, rules, or regularities. In addition, inductive reasoning involves the detection of diversity and irregularities. Klauer & Phye (2008) explained that the inductive reasoning process has the same basic properties that regulate the set of elements, such as classification, analogy, incomplete series, and matrices, which are further used as the type of inductive reasoning tasks.

The concept of inductive reasoning was portrayed as the detection of regularities and the uncovering of irregularities. The components of inductive reasoning consist of several elements that are mapped based on their cause and relation effect. There are six types of inductive reasoning processes described in Table 5. Inductive reasoning tasks can be further categorised in the scope of six different cognitive operations presented in various objects or tuples of objects (verbal, pictorial, geometrical, numerical, and others). The strategy for inductive reasoning tasks lies in the comparison process within attributes or relationships, whether from searching for their similarities or differences. In the educational assessment, the inductive reasoning task was done in different forms of objects. The series and analogy task were found to be one of the most used tasks to assess inductive (Rivera & Becker, 2003).

Table 5. The types of inductive reasoning process (Klauer et al., 2002)

Process	Facet identification	Problem formats	Cognitive operation
Generalization	a1.b1	Class formation and or expansion Finding common attributes	Similarities of attributes
Discrimination	a2.b1	Identifying irregularities	Difference in attributes
Cross-classification	a3.b1	4-, 6-, and or 9-fold scheme	Similarities and difference of attributes
Recognizing relationship	a1.b2	Series completion; analogy	Similarities of relationship
Differentiating relationship	a2.b2	Disturbed series	Different in relationship
System construction	a3.b2	Matrices	Similarities and difference in relationship

The series completion task such as number series and figure series are typical tasks involved in the inductive reasoning test. Generally, series completion task requires the ability to predict the next element which is ordered in a relational pattern. The test taker explores the relationship between elements and finds out what are the next elements in the series (Van Vo & Csapó, 2020). The strategy for solving different type of series task requires various approaches (Liang et al., 2014). The number series task uses retrieval strategy in which the relation between number adjacent elements is attained from the arithmetic fact in the long-term memory. When solving the number series task (e.g. 5 10 15), the test takers directly get the relation by retrieving their arithmetic ability (e.g.  $10 + 5 = 15$ ) and apply it to predict the pattern. The figural series, however, require more complex steps. They are mostly solved with the procedural strategy which require two steps of thinking process, identify the figure complexity (e.g. the shape: rectangulars, dots, circles, and triangles) and predict the relation between the elements based on the pattern series (Liang et al., 2014).

The analogy tasks are related to the ability to perceive knowledge of an event or pattern and apply it to a similar situation (Christie & Gentner, 2014). It requires the skills to grasp the similarities based on the relation of set of events (Holyoak, 2012). This ability is widely used in the intelligence test and supports the problem-solving skills (Hayes et al., 2015; Kunda et al., 2013; Van Der Meer et al., 2008) and useful for investigating the mechanism of reasoning. The analogy task can be expressed in a variety of forms such as numerical or figural. In the analogy task, the most used pattern is the 'generative problems' when the test taker is presented with an incomplete structure and is asked to complete the analogy by generating the missing term (Ichien et al., 2020).

### **2.2.2 The role of inductive reasoning in problem solving**

The reasoning process was believed to be one of the cognitive factors that influence the acquisition of problem solving. Inductive reasoning contributes to information processing during the problem-solving process (Hamers et al., 1998). Inductive reasoning was also described as one of the intelligent components that potentially has a connection to problem solving (Wüstenberg et al., 2012). Intelligence involves reasoning activities that are described as the process of concluding order to achieve goals, thus informing problem-solving and decision-making behaviour (Leighton & Sternberg, 2004). The importance of inductive reasoning in problem-solving is further explained by its ability to facilitate the transfer of past learning experience (Haverty et al., 2000). Through this process, individuals can induce conditions to apply problem-solving strategies effectively. Furthermore, inductive reasoning is

also characterized by representation and generalizations, which significantly support individuals in generating better solutions (Magiera, 2012).

Several studies revealed the positive correlation between inductive reasoning and problem-solving ability (Kyllonen & Kell, 2017; Stadler et al., 2015). Molnár et al., (2013) found a medium correlation between inductive reasoning and domain-general and domain-specific problem-solving ( $r = .43$  and  $.44$ ,  $p < .01$ , respectively). Similarly, a comprehensive study by Wüstenberg et al. (2012) found a high effect of inductive reasoning in explaining the rule knowledge of complex problem-solving operations ( $\beta = .63$ ,  $p < .01$ ), while Wu & Molnár (2018) found a moderate effect of inductive reasoning on problem solving ( $\beta = .24$ ,  $p < .01$ ). They explained that inductive reasoning involves cognitive processes such as comparing objects and attributes, finding similarities and dissimilarities, and generating rules that serve as the components of problem-solving. It plays a significant role in applying strategy to acquire knowledge as the initial stage in solving complex problem. Lotz et al. (2022) support this by explaining that the connection between reasoning and problem solving can be detected in the problem-solving mechanism. As an example, in CPS, problem solvers can isolate one variable that causes the effect. In this part, inductive reasoning ability will help identify relations among variables and build a mental model to complete the task (Lotz et al., 2016). In the other CPS task, the knowledge acquisition phase, reasoning enables problem solvers to select, apply, and review strategies to recognise patterns and rules (Veenman et al., 2004). This phase also enabling students to observe situations, draw inferences, and identify causes to reach unexpected findings, which employ several cognitive strategies such as causal, deductive, and inductive reasoning (Dunbar, 2000).

### **2.2.3 Role of gender in inductive reasoning**

The study of gender differences in cognitive aspects, including inductive reasoning, has been explored in numerous studies (Blum et al., 2016; Soeharto & Csapó, 2022; De Van Vo & Csapó, 2023; Waschl & Burns, 2020; Wicaksono & Korom, 2022a; Wu & Molnár, 2022). The majority of these studies have focused on comparing performance of males and females in inductive reasoning, many of the results showed a tendency to favour males over females. A meta-analysis conducted by Waschl & Burns (2020) on the effect of gender on inductive reasoning revealed small effects ( $g = .13$ ,  $CI = 0.10-0.17$ ,  $p < .01$ ), indicating that males outperform females. However, they also reported significant heterogeneity, suggesting a high diversity in the magnitude of gender effects. Indeed, the effect of gender differences may vary depending on the context of the study and sociocultural factors. For example, a contrary result in favoring

females in inductive reasoning was found by Soeharto & Csapó (2022), explaining that female students from high school perform better than their male peers.

In addition, other studies have found no difference in inductive reasoning performance between males and females (Díaz-Morales & Escribano, 2013; Kambeyo, 2018; Molnár, 2011). Van Vo & Csapó (2020) conducted studies on inductive reasoning assessment across different school grade, resulting no significant difference between male and female performance. Wicaksono & Korom (2023) also found no significant effect of gender in inductive reasoning ( $\beta = .08, p > .05$ ) among high school students. Furthermore, Mousa & Molnár (2020) implemented computer-based training for inductive reasoning, finding no significant difference in pre-test and post-test result for inductive reasoning on both genders.

#### **2.2.4 Scientific competencies and knowledge**

In science education, scientific competence plays an important role in helping students engage in scientific phenomena and solve science-related issues. Indeed, many educational practices and learning processes try to include scientific knowledge, competency, and literacy as one of their main goals for students because students' proficiency in certain topics such as science often affects their future success (Rittle-Johnson, 2017). The importance of scientific concepts and knowledge in science education is recognised by numerous studies (Chang et al., 2010). To improve the scientific competency of students, many teaching practices adapted the knowledge of the content and its integration into the learning process of students (Yeh et al., 2015). The curriculum and syllabus based on content knowledge purposes to enhance scientific literacy are progressively developed (Bankel et al., 2003). PISA categorised three types of knowledge in the science domain: content, procedural, and epistemic. *Content knowledge* is related to the ability to use and recall theories, information, facts, and ideas. *Procedural knowledge* is related to methods and techniques. It is also explained as knowledge of identifying processes or tasks to solve problems (Star, 2005). *Epistemic knowledge* is related to an understanding of the individual's role in justifying the knowledge produced by science (OECD, 2019b). These knowledge domains require the cognitive process to be attained through learning activities. The level of the cognitive process is identified into six domains such as *remember* (ability to recognise and recall the information), *understand* (interpreting, classifying, and summarising information), *apply* (ability to execute and implement the concept), *analyse* (differentiating and organising), *evaluate* (checking and critiquing), and *create* (generating, planning, and producing matters) (Anderson & Krathwohl, 2001). The use of these types and dimensions of knowledge in science topics is the basic core of scientific competency.

In the assessment process, scientific competency can be measured with different types of frameworks. One of the prominent examples is the scientific literacy framework, which represents the competency to engage in discussion about issues that involve scientific knowledge. In the Programme for International Student Assessment (PISA), scientific literacy is described as 'the *capacity to use scientific knowledge, identify questions, and draw evidence-based conclusions to understand and help make decisions about the natural world and the changes made to it through human activity*' (OECD, 2004, p. 16). The measurement of scientific literacy can be defined by the three competencies including (1) *explaining phenomena scientifically* – recognise, offer, and evaluate explanations for a range of natural and technological phenomena, it requires deep understanding to recall and use theories, ideas, information, and facts. (2) *Evaluation and design of scientific enquiry*: describe and evaluate investigations and propose the solution to address problems scientifically; this competency has a connection with the scientific method, procedures, research, and measurements that emphasise knowledge of common procedures used in science. (3) *Interpret data and evidence scientifically*: analyse and evaluate data, make claims and arguments, and draw conclusions (OECD, 2017b, p. 134).

The measurement of scientific competency with the PISA framework has been applied in several countries and the results were used for additional data for specific purposes (Chiang & Tzou, 2018). Scientific competencies align with the various issues and problems that are framed in different situations and contexts of science, such as health, environment, diversity and population dynamic, natural resources, hazards, and frontier of science and technologies (Bybee et al., 2009). In PISA, the evaluation of scientific competence, the situation presented was described at the personal, social, and global levels (Table 6). The scientific competency assessment includes the combination of core competencies, context, domain of cognitive process, and type of knowledge. The interaction of these components was presented in Figure 9.

Table 6. The context of PISA science assessment

	Personal	Local	Global
Health and disease	Maintenance of health, accident, nutrition	Control of disease, food choices, community health	Epidemic, spread of infectious disease
Natural resources	Personal consumption of materials and energy	Maintenance of human populations, quality of life, security, production and distribution of food, energy supply	Renewable and non-renewable natural system, population growth, sustainable use of species
Environmental quality	Environmentally friendly actions, use, and disposal of materials and devices.	Population distribution, disposal of waste, environmental impact	Biodiversity, ecological sustainability, control of pollution, production, and loss of soil/biomass
Hazards	Risk assessment of lifestyle choices	Rapid changes, slow and progressive changes, risk assessment	Climate change, impact of modern communication
Frontiers of science and technologies	Scientific aspects of hobbies, personal technologies, music, and sporting activities	New materials, devices and progresses, genetic modifications, health technology, transport	Extinction of species, exploration of space, origin, and structure of universe.

Note: taken from the PISA 2018 assessment and analytical framework (OECD, 2019a, p. 103).

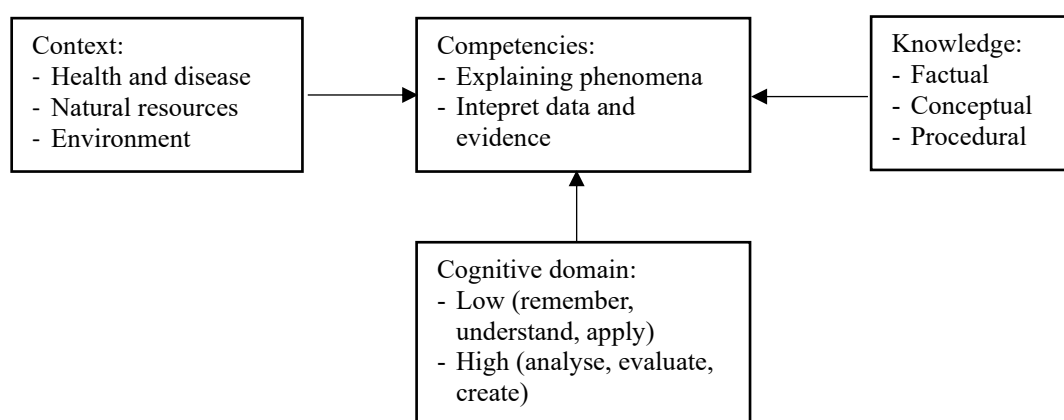


Figure 9. Interrelation between scientific competency components (adapted from OECD, 2019b, p. 102-103)

In some countries including Indonesia, global assessments such as the PISA science assessment are likely constructed with a different context compared to what students learn during teaching and learning practice in school. In Indonesia, the science subject was taught during middle school as an integrated subject. Therefore, it was divided into biology, physics, and chemistry subject in high school. Science assessment was carried out due to teaching needs based on student learning progressions. It was built based on learning competencies and covers several topics to determine student achievement after several stages of learning process (Rahmi

et al., 2021). All assessment was done by teachers, and there is no specific regulation to apply the assessment process. At the end of the school year, there was a national exam to determine whether students could continue their education to the next level. It was a controlled assessment developed by the Ministry of Education that was applied to all schools in Indonesia. The national assessment covers all core competencies from all grades in the schools. As an example, the national exam in secondary schools is conducted at the end of grade 12 and covers all of the core competencies of grade 10 to 12.

In the global assessment, Indonesia has participated in the PISA science assessment since 2003. From 2003 until 2018, the science performance of Indonesian students has fluctuated but remained flat overall, meaning that there has been no significant progress in students' science competencies during last 15 years. It was below the average, for which more than 50% of Indonesian students had a proficiency level below two indicating that students were able to draw a conclusion from a simple dataset and address scientific phenomena in everyday life based on their understanding of content and procedural knowledge. They have difficulty using complex and abstract knowledge and interrelated scientific ideas. They also struggle to discriminate relevant and irrelevant information and evaluate complex experimental design (OECD, 2019c). Therefore, the level of science performance of Indonesian students was quite low compared to other countries, improvement in educational standards and educational equity is urgently needed in Indonesia (OECD, 2019c). In terms of assessment, the standards of science test can be developed based on the national core curriculum and with the influence of global assessment framework.

### **2.2.5 The role of scientific competencies and knowledge in problem solving**

The problem-solving process requires knowledge to understand the problem situation and make conclusions. In domain-general problem-solving assessment, the influence of knowledge and competencies in a specific topic is not fully necessary, rather the test takers will obtain knowledge during the problem-solving activities (Fischer et al., 2012). Content knowledge and competency build a foundation and insight that can help stimulate and support learners to solve problems (Wang et al., 2013). In a specific problem situation, prior knowledge, experience, and skills will contribute to the success of problem-solving activities (Beckmann & Goode, 2017). In addition, Glaser et al. (2009) knowledge and competence in a particular domain will refer to experience that has a benefit for problem solving in the organisation of mental activities to quickly analyse information and effectively describe the representation of their problem. Indeed, an individual with strong competency and knowledge will be able to reduce the complexity of

the problem, resulting in a successful solution (Seifried et al., 2020). The empirical studies in connection between knowledge and problem solving have been done resulting in the strong relationship between conceptual and declarative knowledge to problem solving (Friege & Lind, 2006). Goode & Beckmann (2010) explained that a different level of knowledge and information will influence decision-making in problem solving. In addition, prior knowledge and competency play an important role in problem solving, as Weise et al. (2020) discover the moderating effect of prior knowledge in facilitating the relationship between intelligence and problem solving.

The connection between knowledge and problem solving is further supported by She et al. (2012), who applied web-based learning with problem-solving task. They found that the students with higher domain knowledge in science performed better in problem-solving tasks ( $\beta = .29, p < .01$ ). During the implementation of problem-solving instruction, successful students in solving well-defined problem exhibit higher mastery of knowledge, supporting the positive influence of domain knowledge in problem solving ( $\beta = .19, p < .05$ ) (Yeh et al., 2022). Additionally, Vieira & Tenreiro-Vieira (2016) working on scientific literacy program, explain that competency and knowledge in science are also connected to critical thinking, judgment, decision-making, and problem solving.

### **2.2.6 Role of gender in scientific competency**

The gender differences in scientific competency and general science performance show various findings. The PISA survey found no significant differences between male and female in science literacy in the majority of the participating countries (OECD, 2004). However, males outperforming females was still found in several countries, and vice versa. Therefore, many studies also favour males in science performance (Ganley et al., 2014; Ting-ting Liu & Sun, 2012; Nwosu & Ibe, 2014), while results showing females outperforming males were also found (Japelj Pavešić, 2008; Shahzadi & Nasreen, 2020). Additionally, Reilly et al. (2019) conducted a metaanalysis of gender differences in science performance, showing that in most countries, there are no significant differences between male and females, with an effect size of  $d = -0.95$ . Indeed, there is variability in gender differences, with some countries favouring males and others favouring females. Based on the cognitive theory, males show better visual capabilities, including numerical computation, mechanical, and spatial orientation (Weiss et al., 2003), while females show more advantages in memorization, verbal, and reading tasks (Hyde, 2016). Bell et al., (2006) supporting the differences between males and females cognitive process through

their studies which assessed brain activation during several cognitive tasks. The result shows difference pattern of activation between males and females during cognitive challenges. Moreover, the gender differences in science performance are predicted to be connected to social and environmental factors (Fang & Wei, 2010), including the school programs, curriculum, learning activities, and parental influences (Hyde, 2016; Sung & Hwang, 2013). Indeed, gender equity and cultural perspective also play a role in supporting these phenomena (Cheng et al., 2021; Luu & Freeman, 2011).

## **2.3 The affective factors influencing problem solving**

In addition to cognitive factors, many researchers also emphasize the influence of affective factors in supporting problem-solving processes (Bat Or, 2014; Furinghetti & Morselli, 2009; Guven & Cabakcor, 2013; Shin et al., 2003). The affective factors are said to be intertwined with the cognitive part to shape students' problem-solving performance. Luo et al. (2019) revealed that attitude will help students to feel ready to do problem-solving activities. Moreover, Guven & Cabakcor (2013) found that students' problem-solving attitude and motivational factors are related to problem solving. Indeed, this present study explores and investigates the role of attitude and motivation as potential affective factors in problem solving.

### **2.3.1 Attitude towards science**

Attitude is described as internal states of understanding and feeling that form personal actions (Driscoll & Burner, 2005). When we relate it to 'personal actions', attitude can be translated as a personal tendency to respond to things, people, ideas or places either in a positive or negative way based on prior knowledge and experiences (Oskamp & Schultz, 2005). In this term, attitude is closely related to personal behaviour because attitude involves the desire to respond, the acknowledgement of value, and the acceptance of value (Uluçınar Sağır, 2012). In another interpretation, attitude can be explained as the ability to assess something based on personal judgement and reflection like accepting the fact, rejecting, or even ignoring phenomena. It is also defined as individual judgement and evaluation about something in which there is a degree of favour or disfavour (Reid, 2006). In a broader definition, it can be described as opinions and feelings about something that lead to an action of behaviour as the response.

In a psychological view, attitude is one of the constructs or variables made from observable responses to stimuli with the assumption that its coherence and consistency among those responses (Kerlinger, 1970). Olasehinde & Olatoye (2014) also added that attitude is strongly connected to cognitive construct with emphasis on thinking skills, perceiving, and feeling

toward the cognitive project. Attitude can be divided into implicit and explicit attitudes (Glock & Kovacs, 2013). Implicit attitude is an automatic evaluation that comes from the spontaneous mind whenever the object is present (Jones et al., 2009), which results in sudden reaction and feeling. In contrast, explicit attitude is a thoughtful reaction based on the evaluation of the attitude object (Gawronski & Bodenhausen, 2006). The expression of explicit attitude requires a long process and controlled effort. This attitude makes people evaluate and construct what they have in their minds or retrieve the evaluation from their long-term memory (Bassili & Brown, 2005). This explicit attitude depends on the motivation. Thus, it will be better expressed when people are able to participate in the processes.

The measurement of attitude in the educational context is mostly explained in a particular concept like science, which leads to specific focus studies like attitude towards science. The attitude towards science measurement is still developing based on different settings and improvements. Back then, the attitude toward science is described as a general positive or negative feeling about science, like or dislike science (Crawley & Koballa, 1994). The measurement of attitude towards science developed from a very simple construct like the Attitude Towards Organise Science Scale (ATOSS), which included 4 items, each with two potential response choices 'agree' or 'disagree'. The items talked about science and technology, the benefit of science in life, and science versus faith (Brossard et al., 2005). In the other studies, attitude towards science is explained by some dimensions such as perception about science, anxiety or enjoyment toward science, self-esteem in something related to science, motivation in science, the understanding of the value of science, fear or failure in the science course, the nature of science environment, achievement, and even attitude with peers, friends, and parents (Osborne et al., 2003). Attitude measurement was built over the last decades and has been evolving ever since. With different contexts, changing environments, and discoveries.

The study of attitudes towards science measurement shows diverse constructs with some variation in their dimensions (Navarro et al., 2016; van Aalderen-Smeets & Walma van der Molen, 2013). The attitude towards science is constructed about science and learning science, a sense of scientific experiences, an interest in scientific activities, and working or pursuing a career related to science. Toma & Greca (2018) approach to attitude towards science is composed of the social implications of science, normality of scientists, attitude to scientific inquiry, adoption of scientific attitudes, enjoyment of science lessons, leisure interest in science, and career interest in science. The most prominent and influential model of attitude construct probably is domain-based attitudes which include cognitive, affective, and behavioural dimensions (Dismore & Bailey, 2011; Khine, 2015). The theory proposed by Reid (2006)

classifies attitude towards science into three dimensions: (1) the cognitive dimension, referring to knowledge about science, idea, and belief about science concepts, (2) the affective dimension addressing feelings about science, liking, or disliking science, and (3) the behavioural dimension, covering dispositions towards science activities. This construct was supported by van Aalderen-Smeets and Walma van der Molen (2013) who explained attitude towards science consisting of three dimensions: cognition, affect, and perceived control which is related to behaviour. Kind et al. (2007) also explained that the attitude towards science has clearly defined components such as knowledge about objects, ideas, and beliefs (cognitive), feeling about the object (affective), and tendencies towards action (behavioural). These three domains are independent, yet still closely linked and formed the solid framework of attitudes towards science (Figure 10).

*Cognition.* This dimension refers to cognitive beliefs about science and its impact on human life and society. It is also explained as cognitive judgements which explain the decisions, conclusions, or understanding of science concepts and phenomena (Zhang & Campbell, 2011). The cognitive judgements were raised based on the process of giving value to the science objects. These concepts are strongly related to personal knowledge and experience about science, which indicate the efforts of students in studying science (Schaverien & Cosgrove, 2000). In cognitive dimensions, attitude can be affected by the perception of students about the value and importance of science when it is connected to daily activities and social interactions. Students can identify whether science will give positive and negative value to life and their surroundings or whether it is still relevant to develop science and apply it daily.

*Affect.* The second dimension is connected to emotions, both positive and negative sides of science. The positive emotion was manifested by enjoyment in science, while the negative emotion was represented by anxiety and difficulty with science. Both affective components are not opposite but display independent related subcomponents. The concepts of enjoyment are connected to subjective feelings about pleasure associated with phenomena, events, or particular environments. It arises from the feeling of getting over controlled situations that provide interest and value (Pekrun, 2006). Enjoyment is explained as a positive activated emotion based on focused activities that lead to the joy of experienced success or expected success (Mercan, 2020). In the context of enjoyment towards science, it is related to enthusiasm for science, an inner state of personal feeling about science when they feel good and experience happiness. There are several views about enjoyment being associated with accomplishment as a result of mastering skills, meeting challenges, or satisfaction with performances (Davis et al., 1992; Dismore & Bailey, 2011; Teo & Noyes, 2011). The terms of anxiety come up with the

states of discomfort including fear, tension, worry, distress, frustration, and mental disorganisation (Ma & Xu, 2004). In this context, anxiety is explained in intellectual and personality areas. The intrinsic factors of anxiety involve psychological arousal and negative cognition linked to poor performance, achievement, and the potential for negative evaluation (Ashcraft et al., 2007). This type of anxiety is caused by difficulty during science learning and continues in the personality area leading to a negative feeling or dislike of science.

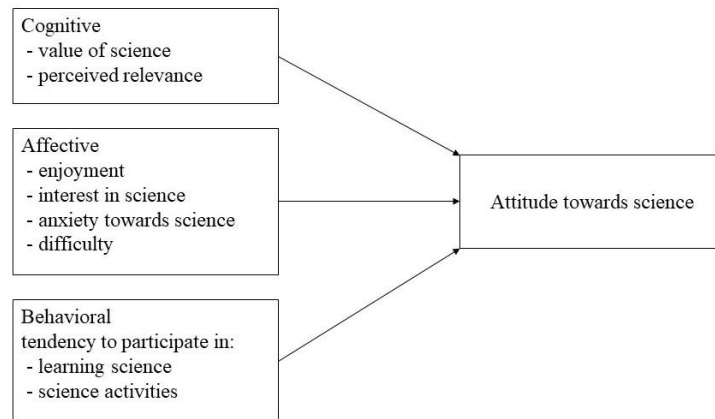


Figure 10. The framework of three dimensions of attitude towards science (adapted from van Aalderen-Smeets et al., 2012)

*Behavior.* Behavioural tendency is a concept that can be used to indices attitude (Mueller, 1986). The behavioural tendency in attitude measurement is pointed to openness in engaging with activities related to science (Zhang & Campbell, 2011). This behavioural and activities are specifically interwoven with learning processes that address formal learning such as classroom learning and informal science learning through outdoor activities and hobbies.

The framework of three dimensions of attitude towards science instruments is focused on three main concepts: (1) whether students appreciate the importance of science in real life and society, (2) whether students are affectively connected to science, express their positive or negative emotions to science, and (3) whether students are involved and willingly to participate in activities related to science or learning behaviour (Zhang & Campbell, 2011). It is expected that students who achieved good attitudes will present all attitudes in those three categories.

### 2.3.2 The role of attitudes in problem solving

Attitude is known to be one of the affective factors that has a great influence on the learning environment (Martin-Dunlop & Fraser, 2007; Taghizadeh & Hajhosseini, 2021). In science education, attitude is one of the most prominent factors that influence learning achievement

(Korur et al., 2016). The importance of attitude toward science is not arguable, and so many educators support its role in developing students' skills through problem-solving activities. Problem-solving strategies that exhibit activities in analysing problems and creating solutions are proven to be able to increase students' attitudes (Ferreira & Trudel, 2012; Gok & Silay, 2010). Problem-solving instruction requires students to collaborate to solve the given problems. Students can develop arguments and ideas and then exchange them in the group, which influences the classroom atmosphere and their attitudes. Two types of arguments that reinforce the application of the attitude toward science during the learning period. The first argument argues that the effective way of learning science is if students act like real scientists in the classroom. It may be obvious that attitude toward science is important in the professional life of scientists and that students who learn science must be aware of how scientists work from simple matter to conclusion. The second claim explains that attitude represents a personal attribute. Regarding the attitude towards science, scientist attributes such as open mind, curiosity, and decision-making are embedded in student problem-solving skills and everyday life (Gauld, 1982).

The problem-solving activities can be enhanced with positive attitude. Some studies revealed a positive correlation between attitude and problem-solving achievement (Güven & Cabakcor, 2013; Marchis, 2013; Öztürk et al., 2020). Luo et al. (2019) found significant effect of attitude to problem-solving ability ( $\beta = .14, p < .05$ ). Furthermore, the influence of attitude in problem solving is presented both as a direct and indirect effect. Since the attitude creates a better learning environment, it can promote students' performance while doing problem-solving tasks. Low attitudes also liked learning difficulty, which increased the chance of having learning difficulty and hindering problem-solving activities. Negative attitudes such as anxiety is often connected to an inability to concentrate, which likely reduces the availability to use cognitive resources (Nolen-Hoeksema, 2000; Owens et al., 2012). The other theory might raise the point that affective factors such as attitude and belief directly influence the choice of proving strategy, reaction to face difficulty in solving problems, and expectation towards the chosen strategy (Furinghetti & Morselli, 2009). Negative attitude is linked to less efficient problem-solving performance (Lai et al., 2015). It creates tension and interferes with the identification process to create a solution.

### **2.3.3 Gender differences in attitude**

The effect of gender concerning attitudes has been extensively studied (Aini et al., 2019; Glock & Kovacs, 2013; Oon et al., 2020; Teo & Noyes, 2011). However, the role of gender in

attitude is still debatable, with some studies finding that males outperformed females in attitude for learning purposes (Hacieminoglu, 2016; Smith et al., 2014). Oon et al. (2020) found that female students' interest in science subjects such as physics is lower than that of male students, with female students also inclined to favor other subjects rather than science. Jones et al. (2000) also reported that male students are more likely to have interest in physics and chemistry subjects (e.g. atom, plane, computer, electricity, and energy), while females are interested in biology-related topic (e.g. health, animal life, and environment), with their overall attitude is significantly different. In contrast, other studies found female students tend to show a more positive attitude compared to male students (Aini et al., 2019; Khishfe & BouJaoude, 2016; Liu et al., 2010; Mihadiz et al., 2011). Aini et al. (2019) found significant different between gender in attitude ( $F = 5.83, p = .015, \eta_p^2 = .004$ ) and explained that female has more awareness that science has beneficial and harmful effect to environment, strengthening their value of science. Furthermore, Susilawati et al. (2022) conducted survey about students' attitudes in Indonesia and found that female students show more awareness about the application of science in daily lives. They are also more likely to show interest in conducting science experiments during learning.

#### **2.3.4 Motivation in science**

In the psychological approach, motivation is commonly derived from the social cognitive theory of human learning (Schumm & Bogner, 2016). In a general term, motivation refers to a feeling of being able to do the duty, an inner force like self-determination and self-efficacy, and a feeling of responsibility (Pintrich & De Groot, 1990). When it is connected to the learning process, motivation is narrowed into the personal internal state that arouses, directs, or sustains their behaviour (Koballa et al., 2007). Motivation covers a self-concept of the ability to respond to specific situations, various tasks, goals, identify personal value, and generate conscious or unconscious choices (Eccles & Wigfield, 2020).

Motivation can be distinguished by two main types: intrinsic and extrinsic. Intrinsic factors refer to any action in the individual that they feel interested in and enjoy, while extrinsic motivation refers to individual involvement in something because they want to achieve specific outcomes. In intrinsic motivation, social context, feeling of inner competence, and autonomy have a strong effect in driving students to get motivated. For example, students who choose to do fieldwork and master that work will be intrinsically more motivated than students who are compelled to do the work. They will show their enthusiasm for working and will engage more in every activity. It is important to remember that to make students experience intrinsic

motivation, they need to get engaged in the learning activity and be satisfied with the process. This 'engagement' is a term that represents the action of individual motivation. Thus, Baron & Corbin (2012) define engagement as: "The engaged student is the student who has a positive, fulfilling and work-related state of mind that is characterised by vigour, dedication, and absorption and who views him or herself as belonging to and an active participant in his or her learning communities." (p. 763). This motivation to engage in activities does not only depend on the external reward. Sometimes people are willing to do something because they feel joy or any other feeling. In some cases, this kind of feeling will drive people to participate in work. It also gives a sense of progress when they see that the work is done and gives a positive result by acquiring new knowledge, skills, or experiences. Even though intrinsic motivation is not related to reward or final result, some factors can increase intrinsic motivation, and some of the factors are feeling of challenge, curiosity, control, competition, and recognition.

In addition, external motivation is related to external rewards, such as a reward or performing tasks required to get a good grade on an assignment. In contrast to intrinsic motivation, this extrinsic motivation is said to be inherently nonautonomous (Abeysekera & Dawson, 2015). For example, a student tries hard to do their homework and understand their subject because they know that it benefits their academic qualification which is needed for a specific job within a specific career. In contrast, another student does assignments and works based on the instruction of his teacher. Both behaviours show external motivation associated with homework rather than any challenge or feeling of satisfaction. In comparison with internal motivation, external motivation tends to be more effective in making students engage in the learning process, for example. Providing interesting rewards will influence people's motivation in a fast way because they want to get the reward. However, offering excessive rewards can lead to a decrease in intrinsic motivation. That means that the activity is only focused on getting rewards without feeling like enjoying and developing the inner feeling of doing certain work.

In the context of science education, motivation is described as an internal state that arouses, directs, and sustains science-learning behaviour (Glynn et al., 2011). Science motivation has a goal to make students enjoy science subjects, recognise the world through science, get more confidence by learning science, and be willing to study and learn more about science. That is why motivation in the science context is one of the most important goals. In a practical context, science motivation is explained in four main features (Glynn et al., 2009).

Self-determination is understood as a consideration of the innate psychological need for individual competency, relatedness, and autonomy (Deci & Ryan, 2000). Individuals with strong self-determination feel competent for doing challenges, confidence in making choices,

and feel connected with others. Self-determination is also defined as autonomy, the need for an individual to act and make real behaviour that represents the feeling of freedom. The need for autonomy comes from the notion of locus of causality or being original about the action rather than being pulled and pushed by external forces. For example, if students are willing to have a break but feel forced to keep learning in the classroom and making some discussion, their autonomy will be thwarted. Self-efficacy is conceptualised as a judgment of internal capability to complete specific tasks and achieve goals (Bandura, 1993). It also refers to an ability to organise and execute action for desired outcomes and is connected to the belief of the students that they can achieve well in science. Students with a high level of self-efficacy can determine which learning strategy to choose based on their previous skills and knowledge. It is also related the ability to organise and execute action for desired outcomes (Hoffman & Schraw, 2009). Intrinsic motivation refers to any action in the individual that they feel interested in and enjoy. It is connected to personal satisfaction with performance rather than the result (Deci & Ryan, 2000). Intrinsic motivation comprises two bases for competence and autonomy that represent feeling and causality (Pedrero & Manzi, 2020). In terms of grade and career motivation, they are considered as external motivation. It acts as mutual supporting components of motivation that positively contribute to directing students' learning behaviour (Glynn et al. 2011).

### **2.3.5 The role of motivation in problem solving**

Many studies prove that motivation will influence performance in doing a task (Cerasoli et al., 2014; Fredricks et al., 2004; Moore et al., 2014; Vollmeyer & Rheinberg, 2000; Yang & Wu, 2012). Student motivation can facilitate or help students during the learning process in problem-solving tasks in a self-regulated way (Baars et al., 2017). The combination of motivation and cognitive or intellectual factors is good at predicting problem solving (Vollmer & Kaufmann, 1975). van Harsel et al. (2019) emphasised the self-efficacy aspect of motivation that students with higher self-efficacy will benefit from engaging in problem-solving performance. Efficacy belief is related to involvement in the task and significantly influences academic achievement in mathematics. The level of self-efficacy is somehow related to intrinsic motivation that influences the student's persistence in a learning task and influences their ability and effort to develop problem-solving ability (Song & Grabowski, 2006). Motivation also encourages the student to engage in problem-solving tasks with a focus on reaching goals (Bat Or, 2014). Specifically, in the self-determination theory of motivation, students use more effort when they find learning materials interesting and bring them satisfaction. In this case, they will place an interest associated with problem-solving performance and better comprehension (Baars et al.,

2017). Self-determination theory divides motivation into intrinsic and extrinsic aspects which both can affect achievement and performance related to the learning goals (Cassidy & Giles, 2009). In mastery goal motivation, the students under mastery goals performed better and put in a lot of effort and strategy while solving the problems (Bereby-Meyer & Kaplan, 2005). In a specific case, motivation is strongly related to learning mathematics and is classified as a key factor in progressing learning (Gasco Txabarri & Villarroel Villamor, 2017). Specific topics need a lot of effort and persistence to achieve better understanding and comprehension. In this sense, motivation was claimed to be more specific in certain domains than in general motivation. High motivation is related to a high score in mathematics achievement and is negatively related to anxiety in mathematics studies. Moreover, in medical studies, motivation plays an important role in learning and academic performance (Kusurkar et al., 2013).

The connection between motivation and problem solving has been studied. Özcan (2016) found a connection between motivation and problem solving, claiming that 13% of problem-solving score is explained by internal motivation, while willingness to do learning activities explain 7% ( $R = .45$ ;  $F = 15.58$ ;  $p < .001$ ). Furthermore, several studies showing that the strength of connection between motivation and problem-solving is fluctuated but seems flat over time. The connection between motivation and problem solving was high in early education. Martin (2009) argues that the younger students entered the school, the higher level of motivation, performance, and engagement compared to late students. MacIntyre et al. (2002) found that the students showed different mean motivation in different grades with the highest mean score in lower grades. Another study also revealed that intrinsic and extrinsic motivation tends to decrease along the grades (Lepper et al., 2005). Student motivation is likely to drop during their time learning at school. In contrast to motivation, problem-solving ability tends to increase over the grade. The ability to make a solution from a problem is significantly different between younger and older students, and older students got a higher result in problem solving (Swanson, 2004). As an example, in Grades 2 to 4, the student problem-solving accuracy is increasing simultaneously (Zheng et al., 2011). The contrasting development of motivation and problem-solving ability across the grades will make the correlation of both variables weaker because the increasing value of one is not administered by the others. However, at the university level, the correlation between motivation and problem solving is getting higher. Therefore, it is assumed that the condition at each school level plays a role in affecting the relationship between motivation and problem-solving ability.

### **2.3.6 Gender differences in motivation**

In motivational studies, research on gender has been conducted by many researchers, emphasizing the importance of individual differences in motivation (Cabras et al., 2023; Debacker & Nelson, 2000; Oppermann et al., 2021; Schürmann & Quaiser-Pohl, 2022). Daher et al. (2021) conducted research with a high school sample and found a significant difference between male and female students in science motivation ( $t = -6.29, p < .001$ ), especially for intrinsic motivation, where gender shows a potential mediation ( $\beta = -.40, CI = -0.322, -0.016$ ). They found that female students excel in intrinsic motivation, self-efficacy, self-determination, grade motivation, and career motivation compared to male students. Furthermore, they explained that female students are more aware of the importance of science to their future career, which makes them realize the urge to learn better in science. Ozbas (2016) also found that female students from general high school and vocational high school exhibit higher internal and external motivation than their male peers. Salmi & Thuneberg (2019) revealed that female students are better in situational motivation than male students, but in terms of self-determination, they found no significant differences between male and female students. Another result comes from Bećirović (2017), who found no difference between male and female students in intrinsic motivation in foreign language learning. Similarly, during the implementation of learning process, female students also showed higher motivational level compared to male students in the beginning. However, after experiencing sequences of similar learning activities, the differences between male and female become lower, suggesting that environmental factors can reduce the impact of gender differences in learning motivation (Chung & Chang, 2017).

### **2.4 The theoretical model of the connection between cognitive and affective factors, gender, and socioeconomic status on problem solving**

Domain-general and domain-specific problem-solving is associated with many cognitive and affective factors in educational psychology. Problem solving itself can be explained either in general or specific domains with each type having different context representations, yet a similar cognitive process. The interconnection between domain-general and domain-specific problem-solving was positive and became greater over time due to the similarity of strategy used in both problem-solving situations (Molnár et al., 2013). Both problem-solving strategies require cognitive activities that involve the ability in knowledge processing and application. To this term, the reasoning process has been said to influence problem-solving achievement. Indeed, the positive effect of inductive reasoning on problem solving has been proven in many studies (Kyllonen & Kell, 2017; Stadler et al., 2015) suggesting that inductive reasoning helps

students recognise the problem-solving mechanism and its variables. Moreover, knowledge and competency also contribute to the problem-solving process. Several studies found that prior knowledge mediates problem-solving abilities (Weise et al., 2020) and helps learners build their understanding of problem representation (Glaser et al., 2009). Thus, in domain-specific problem-solving, the role of knowledge and competency is obvious due to the interconnection of problem situation and specific knowledge requirements.

In affective factors, attitude is considered as variables that able to support students' problem-solving during learning activities and their achievements in general (Güven & Cabakcor, 2013; Marchis, 2013; Öztürk et al., 2020). Along with attitudes, many studies also involved motivation as a powerful affective factor to influence problem-solving achievements (Bereby-Meyer & Kaplan, 2005; Song & Grabowski, 2006). Both attitude and motivation contribute to creating an appropriate learning environment. Students with a sufficient level of attitude and motivation are likely to participate and be involved in problem-solving activities. They will show better performance that drives better results in learning with problem-solving scenarios. Thus, in addition to cognitive and affective factors, it is also important to study problem solving from the perspective of individual differences between students according to gender and socioeconomic status. The role of gender and socioeconomic aspects in the achievement of problem-solving by students was emphasised in many studies (Chiu, 2022; Dindar, 2018; OECD, 2014; Wüstenberg et al., 2014) which provide more information to explain the problem solving. Thus, the present study constructed the theoretical model of the connection between cognitive and affective factors to problem solving, which can be seen in Figure 11.

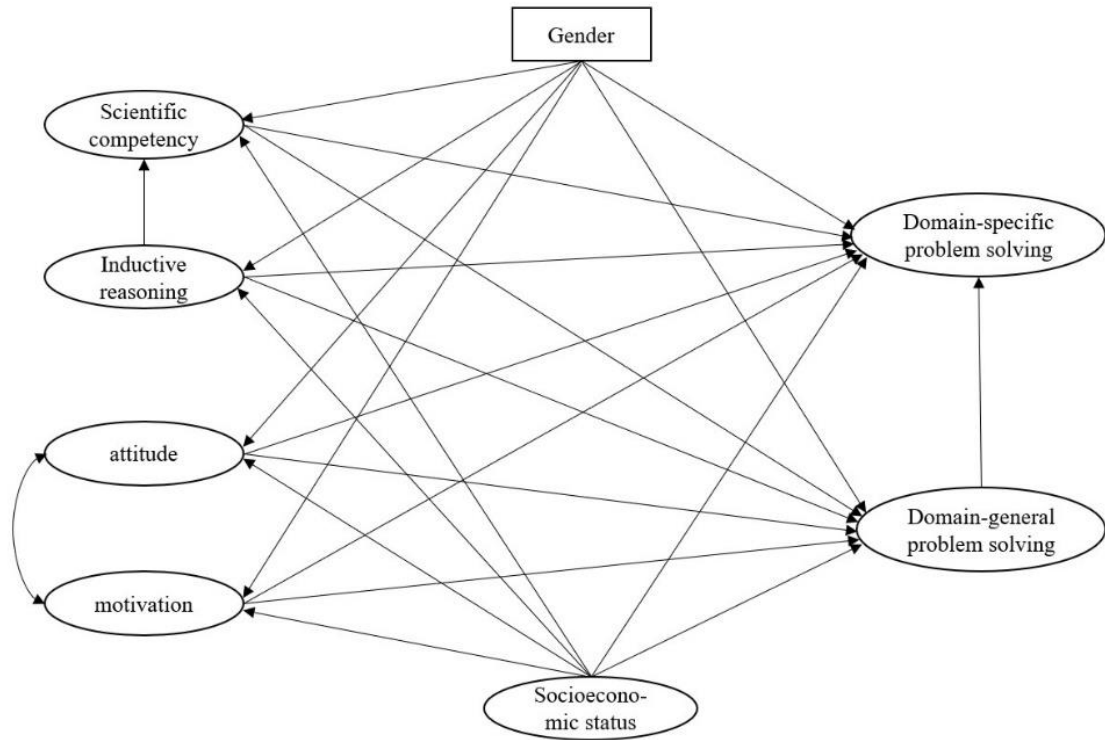


Figure 11. The theoretical model of the relationship between cognitive and affective factors, gender, and socioeconomic status to problem solving

## **CHAPTER 3 RESEARCH AIMS, QUESTIONS, AND HYPOTHESES**

### **3.1 Aims of the research**

This study aimed to measure the problem-solving ability of students in Indonesia. A large-scale survey was conducted to collect information on student problem-solving performance. The students are 10<sup>th</sup>-grade high schooler from various districts in Indonesia. The Indonesian regulation obliged every student to undergo 9 years compulsory education. Indeed, the present study involved students who completed compulsory education and passed national assessment. The measurement of problem solving was categorised into two different domains, general and specific problem solving was measured by two main tests. General problem solving is assessed using the MicroDYN test, while domain-specific problem-solving is assessed with a science problem-solving test.

The other objectives of this study were to investigate the factors that influence problem-solving ability on both the cognitive and affective side. To achieve this, the study involved the measurement of students' ability in cognitive and affective factors, followed by empirical analysis of the relationship between these factors in problem solving. In the cognitive factor, inductive reasoning and scientific competency tests were implemented, whereas in affective factors, we employed attitude towards science and science motivation questionnaires. The adaptation and translation of the tests and questionnaires were carried out, followed by the analysis of the psychometric properties of each test and questionnaire to ensure its validity. Other information on the gender and socioeconomic status of participants is collected as additional factors from the perspective of individual differences to explain the problem-solving ability.

### **3.2 Research questions and hypotheses**

To achieve the research aims, this study formulated research questions and hypotheses as described below.

- RQ 1: What are the students' domain-general and domain-specific problem-solving, inductive reasoning, scientific competence, attitudes towards science and science motivation?
- RQ 2: Is there a significant difference in domain-general and domain-specific problem-solving, inductive reasoning, scientific competence, attitudes towards science and science motivation based on gender?

RQ 3: To what extent do inductive reasoning, scientific competence, attitudes towards science, science motivation, gender, and socioeconomic status influence problem-solving ability?

The hypotheses are formulated based on the theoretical background, with the details described below.

- H1: There is a significant difference between genders in domain-general problem-solving performance, with males outperforming females (Greiff et al., 2018; Wüstenberg et al., 2014).
- H2: There is a significant difference between genders in domain-specific problem-solving performance, with males outperforming females (Gok, 2014; Soto-Ardila et al., 2022).
- H3: There is a significant difference between genders in inductive reasoning performance (Blum et al., 2016), when males performed better in figural and numerical tasks (Waschl & Burns, 2020).
- H4: There is a significant difference between genders in scientific competency, when females outperform males (OECD, 2004; Reilly, 2012).
- H5: There is a significant difference between genders in attitude towards science, when females outperform males (Aini et al., 2019; Liu et al., 2010; Mihadiz et al., 2011).
- H6: There is a significant difference between genders in science motivation, when females outperform males (Oppermann et al., 2021; Schürmann & Quaiser-Pohl, 2022).
- H7: Inductive reasoning is associated with problem solving, meaning that students with higher inductive reasoning ability will perform better in domain-general problem-solving and domain-specific problem-solving tasks (Molnár et al., 2013).
- H8: Scientific competency that represents domain knowledge is associated with problem solving (Glaser et al., 2009; Wang et al., 2013), meaning that students with higher scientific competency will show better performance in domain-general problem-solving and domain-specific problem-solving tasks.
- H9: A positive attitude toward science will lead to higher performance in domain-general and domain-specific problem solving (Güven & Cabakcor, 2013).
- H10: Higher motivation is associated with higher performance in domain-general and domain-specific problem solving (Bat Or, 2014).
- H11: Gender difference is associated with domain-general and domain-specific problem-solving performance (Huang & Chen, 2016).
- H12: Higher socio-economic status is associated with higher performance in domain-general and domain-specific problem solving (Chiu, 2022).

## **CHAPTER 4 RESEARCH METHODOLOGY**

### **4.1 Research design**

The present study used descriptive approach to address the assessment of the problem-solving ability of students in Indonesia. The descriptive approach will characterise several traits based on specific categorisation (Strayer et al., 1995). In this study we explored and classified the level of the student in problem-solving ability, as well as in cognitive factors such as inductive reasoning and scientific competency, and affective factors such as attitude towards science and science motivation. Additional comparative approaches were conducted to compare students' ability levels based on their gender, which shows the trends and patterns of the students' ability distribution.

An associational research design was also implemented to determine the relationship between the measured variables. Associational study consists of predictor or independent variables and outcome or dependent variables that are continuous (Gliner et al., 2016). Based on the theoretical construct, problem-solving ability acted as the dependent variable and cognitive and affective factors as the independent variables. The degree of association of each variable will be measured to determine the strength of the interconnection among variables.

### **4.2 Sample**

The participants in the research were 10th-grade high school students from Java, Indonesia. Clustered random sampling was conducted by randomly selecting schools in the urban area in Java, Indonesia. Fifty-five 10th-grade science classes from twelve schools were selected. All students with approximately 20 to 30 students in each class participated in the survey. The 10<sup>th</sup>-grade students were chosen considering the Indonesian national K-13 curriculum, as this grade marks the transition from integrated science to separate science subjects. Furthermore, students in grade 10 students represent the age group that has completed compulsory education in Indonesia according to national regulations. The students were enrolled in public high schools that followed national curriculum with the same content standard for learning objectives and context through the school programme (BSNP, 2013). A total of 1243 students participated in this study. The detailed background of the participants is presented in Table 7.

Table 7. The characteristics of the participants

Characteristic	N (frequency %)
Gender	
1. Male	457 (36.8%)
2. Female	786 (63.2%)
Age	$M = 16.78$ ( $SD = 0.98$ )
Socio-economic status (SES)	
1. Father education	
a. Elementary school	154 (12.4%)
b. Middle school	117 (9.4%)
c. High school	426 (34.3%)
d. Vocational school	127 (10.2%)
e. University	419 (33.7%)
2. Mother education	
a. Elementary school	145 (11.7%)
b. Middle school	172 (13.8%)
c. High school	430 (34.6%)
d. Vocational school	88 (7.1%)
e. University	403 (32.8%)

### 4.3 Instruments

This study used two different types of measurements, the test to measure domain-general and domain-specific problem-solving, inductive reasoning, and scientific competency, and the questionnaire to measure the attitude towards science, science motivation, and background information of students. All tests and questionnaires were generated in English, and then back and forward translation was done into Indonesian language with the assistance of two reviewers. The instruments were administered via an online platform, the electronic diagnostic assessment system (eDia) (Csapó & Molnár, 2019). The details of the instruments are described below.

#### 4.3.1 Domain-general problem-solving test: MicroDYN

In the domain-general problem-solving assessment, we used the MicroDYN complex problem-solving test, a computer adaptive testing that measures problem-solving ability in two main principles, knowledge acquisition, and knowledge application (Greiff et al., 2015). MicroDYN has been proven valid to measure problem-solving ability from several studies and was included in large-scale assessments (Greiff & Funke, 2017). The MicroDYN test applied a linear structural equation system which allows students to model the system by connecting the input variables and output variables (Funke, 2010). MicroDYN administration starts with instructions on how to operate the system and practice items that make students familiar with the test interface. When facing the test, students will see the system, they freely explore the system by manipulating the value in the output variable (i.e. topax and floba, see Figure 4.1). Students can increase or decrease the amount of input variable (“+” button to increase the value

and "-" to decrease the value) and by clicking the "Apply" button, students execute their exploration activities and will see the changes in the output variables. Then, students can evaluate their exploration activities and identify the connection between input and output variables.

There are two phases in the MicroDYN test known as knowledge acquisition and knowledge application. In the knowledge acquisition phase, the students are presented with the test system that shows input and output variables. They can explore the system and find the connection between the input and output variables. Simultaneously, they can draw the connection model between variables by creating an arrow to connect the variables (Figure 12). In the knowledge application phase, the correct connection between input and output variables is presented in the system. Students are asked to reach a certain value of the output variable (indicated by the red line) by controlling and manipulating the exact value of the input variables (Figure 13). The decision to manipulate the exogenous variable requires careful analysis because the increase of output variable would lead to an increase in one or more output variables, or even a decrease, neutral, or mixed effect (Greiff & Funke, 2017). The actions to test the knowledge application phase were limited by the time and amount of four attempts to the process.

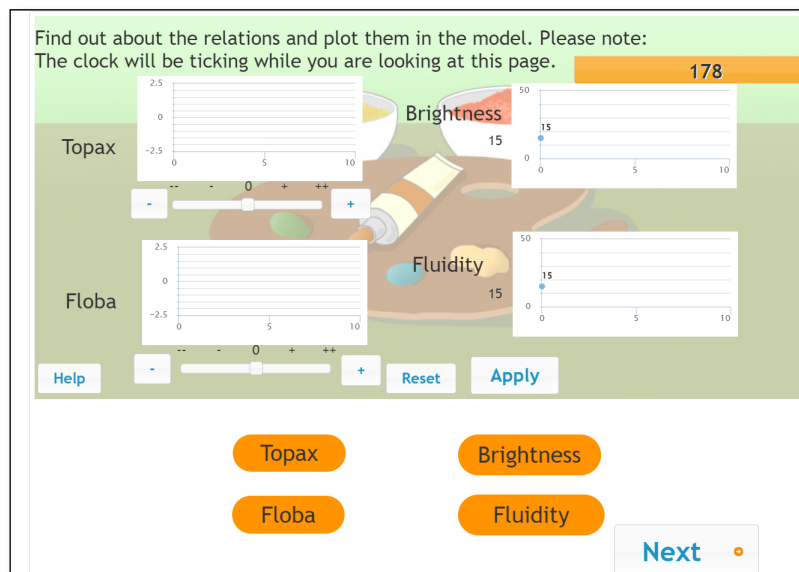


Figure 12. The knowledge exploration phase. The student has to explore the effect of topax and floba on the exogenous variables and continue with drawing line to connect the related variables.

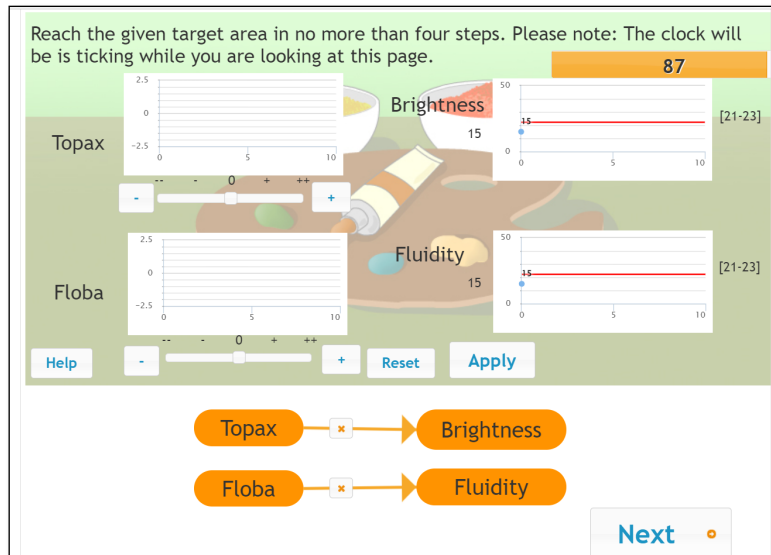


Figure 13. The knowledge acquisition phase. Students must reach the expected value (red line) of the endogenous variable by adjusting the value of the exogenous variable.

The MicroDYN test in this study consists of 20 items, with 10 items in each knowledge phase (see Appendix A). All items showed different themes and complexity indicated by the number of variables presented in each test item and its interactions (Table 8). The type of scoring of the MicroDYN test followed dichotomous scoring when a score 1 was given to students who correctly addressed the test items and 0 was given to students who did not provide the correct response.

Table 8. The complexity of the MicroDYN test

Variable	Complexity level	Number of input and output variables	Number of connections	Items
Knowledge acquisition	simple	(2 or 3) - 2	2 or 3	1, 3, 5
	complex	3 - 3	3 or 4	7, 9, 11
	eigendynamic	3 - (2 or 3)	(2 or 3) + 1 eigendynamic	13, 15, 17, 19
Knowledge application	simple	(2 or 3) - 2	2 or 3	2, 4, 6
	complex	3 - 3	3 or 4	8, 10, 12
	eigendynamic	3 - (2 or 3)	(2 or 3) + 1 eigendynamic	14, 16, 18, 20

#### 4.3.2 Domain-specific problem-solving test

The science problem-solving (SPS) test exhibits the problem-solving measurement in the specific domain. The test framework was derived from the PISA framework and Pólya's problem solving, which focused on the ability to identify problem situation and generate the

solution (OECD, 2016; Pólya, 1945). A total of 11 items were used to measure the science problem-solving ability which was constructed in a multiple-choice test (see Figure 14). Among the test items, five questions represent the first category of problem solving, *identifying the problem* (IP) that assesses students' performance to identify components or variables related to the problem situation and find its relationship to conclude the main cause of problems. The second category, the *generating solution* (GS), consists of 6 items. It assesses the students' ability to analyse information and choose the best solution to address problems. Each item in the test presented different situations in the scope of science contexts such as pollution, bioremediation, animal behaviour, agriculture, pest and disease, conservation and animal protection, and health. The detailed description of each science problem-solving item is presented in Table 9.

Table 9. The description context of science problem-solving test

Category	Code	Description
Identifying problem	IP1	Identification of heavy metal pollution in rivers. The students identify the physical, chemical, and biological indicators of heavy metal pollution and determine the critical amount of each indicator to classify the level of pollution.
	IP2	Identification of factors that influence bird migration in forest and tree formations (shading, leaf petiole, and foraging index).
	IP3	Identification the most potential species that cause fungal disease and attack tomato plants based on their symptoms and physical environment.
	IP4	Determining the cause of microplastic pollution and other contaminants in the pristine areas
	IP5	Identification the level of microplastic pollution based on its residual level found in the coral reef ecosystem.
Generating solution	GS1	Determining the rate of bioremediation to test the effectiveness of microbial agents.
	GS2	Deforestation and tree formation to assist in bird migration.
	GS3	Determining parental plants for crossbreeding to produce high-resistance plants against fungal infections.
	GS4	Creating an exercise plan for the diet based on daily calorie intake.
	GS5	The conservation program to prevent the endanger of giant salamander in the rocky mountain
	GS6	Determine the best conservation location based on the characteristic of the natural habitat of the animal.

The chosen topic was aligned with the Indonesian K-13 national curriculum and was taught to students from middle school to secondary school. The test was designed in multiple choice form followed by the adaptation from several standardised domain-specific problem-solving tests in science (Nadapdap & Istiyono, 2017; Szeberényi, 2013). The test is subjected to content check by two reviewers who are experts in the science and biology fields. The scoring process assigned 1 point to students who correctly answer the questions and 0 point to incorrect answers.

**Plant crossbreeding**  
 The scientists proved that the hybrid production of the new plant's varieties of tomato can improve their resistance to fungal disease (i.e., powdery mildew) and improve the quality of the fruit production. There are several parental tomato plants that have potential to be involved in inbreeding process. In this case big fruit, long stem, and intolerant traits are dominant over small fruit, short stem, and resistance trait (Table 4).

Table 4. The traits of different tomato plants

Parental	Attributes (Traits)
TM1	Big fruit, long stem, intolerant to powdery mildew
TM2	Small fruit, long stem, intolerant to powdery mildew
TM3	Big fruit, long stem, resistance to powdery mildew
TM4	Small fruit, long stem, resistance to powdery mildew
TM5	Big fruit, short stem, intolerant to powdery mildew
TM6	Small fruit, short stem, intolerant to powdery mildew
TM7	Big fruit, short stem, resistance to powdery mildew
TM8	Small fruit, short stem, resistance to powdery mildew

6. Which parental plant that suitable for producing tomato that has big fruit, short stem, and high resistance of fungal disease?  
 a. TM1 X TM3  
 b. TM2 X TM4  
 c. TM5 X TM6  
 d. TM3 X TM7  
 e. TM7 X TM8

Figure 14. An example of a science-problem solving test in the generating solution phase. The students facing a problem related to pests and tomato plants. The students are asked to formulate a solution by selecting suitable parental plants with specific phenotypic traits for crossbreeding to enhance the offspring plants' resistance to pests.

### 4.3.3 Inductive reasoning test

The test to measure inductive reasoning ability in this study was focused on recognising the relationship between attributes by detecting their similarities and dissimilarity. The students will detect similarities and dissimilarities between attributes, find their relationship, and predict the next attribute compatible with the following pattern. We used figure and number attributes for inductive reasoning measurement to represent general symbols that can be used in many contexts. Thus, we used four different test formats including figure series (FS), figure analogy (FA), number analogy (NA), and number series (NS)(Csapó, 1997; Pásztor et al., 2017). The test consists of 32 items, each task composed of 8 items (see Appendix C). The scoring process for each test item was assigned 1 point for the correct answer and 0 point for the incorrect answer. Some examples from the inductive reasoning test are shown in Figure 15.

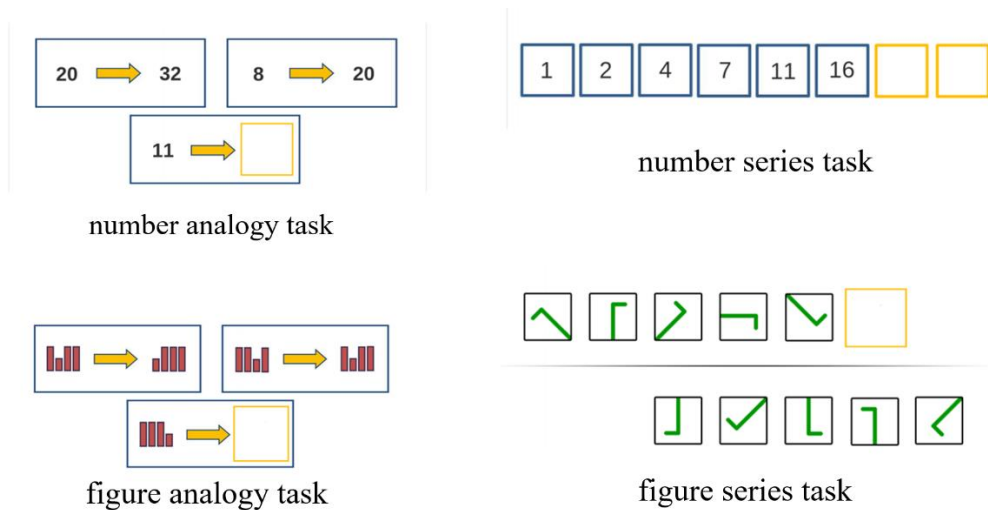


Figure 15. The inductive reasoning tasks

#### 4.3.4 Scientific competency test

In this test, we conducted a task focused on assessing cognitive disposition to deal with science issues at a certain level of cognitive domains and knowledge. The scientific competency test framework work was adapted from PISA scientific literacy, which focused on two main sub-competencies (OECD, 2019b). The first competency, *explaining scientific phenomena*, measures the students' understanding of factual and conceptual knowledge from a certain science topic represented by 8 items. The second competency, *interpreting and evaluating scientific data or evidence* measures students' ability to analyse and evaluate factual and procedural knowledge related to specific science issues and contains 7 items (see Figure 16). Unlike science problem-solving tests, which present problem situations to solve, the scientific competency test aims to measure the understanding of scientific concepts based on prior knowledge and comprehension of new information. A total of 15 multiple-choice items were generated for measure scientific competency (see Appendix D). When the students completed the test, the correct answer was assigned as 1 point, and the incorrect answer was assigned as 0 point. The details of the scientific competency test description are shown in Table 10.

Table 10. Description of the scientific competency test

Competency	Items no.	Science topic
Explaining scientific phenomena (SCE)	1	symbiotic of mycorrhiza and legumes
	2	inorganic compound and legumes development
	4	external factor in bird migration
	7	the effect of biogas application
	9	bioremediation process
	11	the cause of global warming
	12	food chain
Interpreting and evaluating scientific data or evidence (SCI)	15	organism interaction
	3	effect of organic compound and mycorrhiza on the development of legumes
	5	migration period
	6	effect of temperature on biogas production
	8	accumulation of heavy metal in plants
	10	the progression of global warming
	13	balance of the ecosystem
14	biological control of pests	

9. Explain how fungal strains help in the chromium remediation process?

  - a. absorbing the pollutant and accumulate it in their body.
  - b. mobilising the chemical compound into their host plants;
  - c. change the pH of their surrounding with the pollutant compound
  - d. degrading the pollutant compound to a less toxic form
  - e. produce mucus to trap the pollutant.

Figure 16. An example from the scientific competency test. The question is focused on assessing students' understanding of science concepts and explaining the role of microorganism in pollution reduction.

#### 4.3.5 Attitudes towards science questionnaire

The attitude towards science was measured by an adapted questionnaire from several studies (Kind et al., 2007; van Aalderen-Smeets et al., 2012; Wendt & Rockinson-Szapkiw, 2018). We took several items from the previous questionnaire and adapted them based on the Indonesian context. The rearrangement of the items in the attitudes towards science questionnaire (ATSQ) was performed according to the three-dimensional framework and grouped into four latent variables. Firstly, the affective dimension expresses the emotions to science which were described into two latent variables, enjoyment and confidence as positive emotions and anxiety and difficulty as negative emotions (e.g., “*I feel happy while learning science*” and “*I feel nervous while learning science*” respectively). Second, the cognitive domain expressed the

cognitive belief about the value of science (e.g., “*The benefits of science are greater than the harmful effects*”). The last behavioural domain expresses the tendency to participate in learning and activities (e.g., “*I like to participate in science class discussion*”). The total item for attitude towards science measurement was 27 items, divided into four latent variables, which enjoyment variable was explained by 8 items, anxiety and difficulty consist of 7 items, participation in science learning and activities with 7 items, and the value of science consist of 4 items (see Appendix E). The content of the ATSQ was verified by two reviewers to ensure that the items are suitable for measuring attitude toward science in an Indonesian context. The questionnaire is presented in an online version with the scoring type followed by a Likert scale ranging from 1 to 5 (1 = strongly disagree, 2 = disagree, 3 = neutral, rather agree or disagree, 4 = agree, and 5 = strongly agree).

#### **4.3.6 Science motivation questionnaire**

The science motivation questionnaire used in this study was adapted from SMQ II (Glynn et al., 2009, 2011). The questionnaire assesses motivation related to learning science which expresses students’ science learning behaviour. The assessment was explained in 5 latent variables, including intrinsic motivation, career motivation, self-determination, self-efficacy, and grade motivation. First, in the intrinsic motivation variable, the students make self-assessments about their interest in learning science that benefits for its own sake (Glynn et al., 2009) (e.g. *I am curious about discoveries in science*). Second, career motivation, it assessed students' motivation to learn science for future career benefit (e.g. *My career will involve science*). Third, self-determination that describes student motivation based on their psychological needs (Ryan & Deci, 2000) (e.g. *I spend a lot of time learning science*). Fourth, self-efficacy, it assessed students’ belief in their capabilities to learn science (Lawson et al., 2007) (e.g. *I believe I can master science knowledge and skills*). And fifth, grade motivation focuses on external motivation for learning science to get better grades (e.g. *Getting a good science grade is important to me*). Each variable consists of 5 items, with a total of 25 items generated to measure students’ science motivation (see Appendix F). The students responded to this questionnaire choosing the options based on the Likert scale ranging from 1 to 5 (1 = strongly disagree, 2 = disagree, 3 = neutral, rather agree or disagree, 4 = agree, and 5 = strongly agree).

### **4.3.7 Background questionnaire**

The background questionnaire was applied in this study to get additional information about students' gender and socio-economic status (SES). The background questionnaire measures the identification data of the students about gender, age, and parent's education (see Table 4.1). In SES, students input information about their father and mother's highest education level. The level was placed into five categories, 1 for elementary schools, 2 for middle school, 3 for secondary school, 4 for vocational high school, and 5 for university or higher education level.

### **4.4 Procedures**

The research instrument was administered on the online eDia platform (Csapó & Molnár, 2019), and each test and questionnaire was stored in separate domains. Each instrument was sent to students through the teachers. We contacted the teachers in the selected school sample and asked them to distribute the online instruments to students during school time. The students voluntarily participated in the study and consent was obtained before the distribution of the tests and questionnaires. The students took the test and filled out the questionnaire in the school laboratory under the supervision of the teacher. Students were given specific time to complete the instruments in three consecutive phases. Firstly, they did the background questionnaire, attitude towards science, and science motivation, they spent approximately 60 minutes completing the questionnaires. Secondly, the scientific competency and inductive reasoning test were given, and students completed in 60 minutes each. Lastly, the domain-general problem-solving task was given to the students for 90 minutes completion, and it is followed by the domain-specific problem-solving test for 60 minutes. After students complete the tests and questionnaires, they get automated responses about their performance. The exclusion criteria applied when student did not participate in one of the tests or questionnaires. Additionally, incomplete responses in each test and questionnaire are excluded from the analysis. All the data was collected on the eDia server and are kept confidential due to information protection.

### **4.5 Data analysis**

Data were subjected to quantitative analysis to address research objectives. First, the validation of the instruments was presented with a focus on construct validity and reliability analysis. Confirmatory factor analysis (CFA) was used to ensure the construct of the tests based on the theoretical framework. We apply the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of the approximation (RMSEA), and the standardised root mean square residual (SRMR) as a parameter to describe the model with cutoff value as follows:

CFI > .90; TLI > .90; RMSEA < .08; SRMR < .08 (Hu & Bentler, 1999). The factor loading of each item was measured with a value greater than 0.40 for an acceptable result. Furthermore, multidimensional Rasch model analysis was also performed to verify psychometry properties and quality of the test items (Boone et al., 2014). The Rasch model constructs a linear measure logit scale based on the difficulty and probability of the items for the participants to be able to answer the questions (Sbeglia & Nehm, 2019). The infit mean square (MNSQ) value is used to determine the item fit, with an acceptable range between 0.5 and 1.5 (Boone et al., 2014). The bias detection of the test was performed using different item functioning (DIF) analysis with respect to gender groups. The item bias was indicated by a significant result of DIF, whereas a nonsignificant result will confirm no bias identified. In addition to that, the size of the DIF effect was calculated to quantify the bias effect, where the low size (< .64) can be interpreted as negligible (Boone et al., 2014). Reliability analysis (Cronbach's alpha and McDonald's omega) was also added to confirm the internal consistency of the instruments.

The descriptive analysis was performed to profile the students' ability and their level in each variable measure. We used logit value for standardized measurement to describe students' ability levels. This logit value was generated from Rasch's analysis representing the level of student ability in each test and questionnaire responses based on standardization estimates (Boone, 2016). The comparison analysis was conducted using the t-test to compare the level of student ability according to their gender. To answer the connection between the cognitive and affective factors and problem-solving ability, we performed a structural equation model (SEM) with a maximum likelihood (ML) estimator to analyse the relationship between observed and latent variables (Khine, 2013). The data analysis process was performed with SPSS software for the profile and comparison analysis of the students' abilities. The Winstep programme was used for the Rasch analysis. Thus, the MPLUS software was used to perform SEM and factor analysis.

## CHAPTER 5 EMPIRICAL RESULTS

In this chapter, the results of the empirical study are presented. There are seven subchapters explained from 5.1 to 5.6, focusing on the psychometric properties of the measures and the description of students' abilities in each trait and a comparison of students' abilities based on gender differences. Subchapter 5.7 focuses on the relationship between cognitive and affective factors, gender, and socio-economic status in domain-general and domain-specific problem solving. I included several published studies for this chapter (Wicaksono & Korom, 2022a, 2023a, 2023b, 2021, 2022c, 2023c) with detailed explanations of each subchapter are presented below.

### 5.1 The assessment of domain-general problem-solving

We used the MicroDYN test to assess domain-general problem-solving based on the CPS approach. The test has two phases which include knowledge acquisition and knowledge application tasks, which serve as the foundation of the problem-solving process. To ensure the validity of the domain-general problem-solving test in an Indonesian context, the psychometric properties of MicroDYN were examined before the assessment process. Additionally, a descriptive analysis was conducted to profile students' domain-general problem-solving abilities, including a comparison between male and female students. The first hypothesis (H1) was tested in this subchapter, predicting that there is significant difference among gender with males outperforming females in domain-general problem solving (Greiff et al., 2018; Wüstenberg et al., 2014).

#### 5.1.1 The validation of MicroDYN test

The MicroDYN test shows good reliability coefficients ( $\alpha = .849$  and  $\omega = .859$ ) confirming that the test shows internal consistency. The 2-factor CFA was performed to ensure the construct validity of the test which resulted in an acceptable model fit with  $\chi^2/df = 1.98$ , CFI = .99, TLI = .99, RMSEA = .03, and SRMR = .08. The factor loading of each item in the knowledge acquisition phase is ranged from 0.58 – 0.97 while in the knowledge application is also ranged from 0.57 – 0.90, showing that the items belong to each construct.

Further analysis with the Rasch model was performed to measure the quality of the individual items based on the fitting index. We performed an unidimensionality and multidimensionality analysis (2 dimensions) to check the fit index of the items (Table 11). Among the 20 elements of the MicroDYN test, all elements have a good fit mean square

(MNSQ) value ranging from 0.67 – 1.36 which means that each element has good quality to measure the domain-general problem-solving. Based on a two-dimension model, the items have a good fit MNSQ (0.70 – 1.29) that supports the theoretical construct of having two dimensions of the MicroDYN test. The other parameter to ensure validity is the correlation of the item's point measure (PTMA) that shows a positive value.

Regarding the difficulty level of the test, in knowledge acquisition, item 18 (logit measure = 2.88) is categorised as the most difficult item, while item 1 (logit measure = -3.58) is the easiest. This difficulty level of each item was aligned with the task structure where items 1 and 2 were constructed with the less variable and simple linear structure (2x2) whereas items 13 and 18 have a more variable and complex structure (3x3). The Wright map of MicroDYN test items is presented in Figure 17.

Table 11. The item fit of the MicroDYN test from Rasch analysis

Variable	Item	Total score	unidimension			2-dimensions			PTMA
			Measure	SE	Infit MNSQ	Measure	SE	Infit MNSQ	
Knowledge acquisition	d01	479	-3.58	0.08	1.21	-2.99	0.07	1.29	0.57
	d03	216	-1.67	0.10	0.84	-1.10	0.09	0.89	0.66
	d05	370	-2.88	0.08	1.17	-2.32	0.08	1.22	0.58
	d07	343	-2.70	0.08	0.88	-2.13	0.08	0.98	0.68
	d09	190	-1.41	0.10	0.67	-0.83	0.10	0.70	0.72
	d11	222	-1.73	0.10	0.67	-1.15	0.09	0.71	0.73
	d13	20	2.25	0.26	0.73	3.00	0.19	1.01	0.37
	d15	33	1.57	0.21	0.85	2.55	0.17	1.03	0.42
	d17	30	1.69	0.22	0.82	2.40	0.17	0.95	0.40
d19	22	2.12	0.25	0.73	2.87	0.37	0.95	0.38	
Knowledge application	d02	288	-2.29	0.09	1.28	-2.68	0.08	1.10	0.54
	d04	256	-2.03	0.09	1.26	-2.44	0.08	1.05	0.54
	d06	56	0.76	0.17	1.01	0.10	0.13	0.91	0.44
	d08	26	1.89	0.23	0.90	1.14	0.16	0.83	0.38
	d10	86	0.05	0.14	0.87	-0.55	0.11	0.84	0.55
	d12	65	0.52	0.16	0.96	-1.12	0.12	0.90	0.50
	d14	60	0.65	0.16	1.10	-0.00	0.13	1.03	0.44
	d16	40	1.27	0.19	0.96	0.57	0.14	0.95	0.43
	d18	6	3.72	0.43	1.14	2.88	0.19	1.02	0.14
d20	27	1.84	0.23	1.36	1.10	0.40	1.18	0.26	

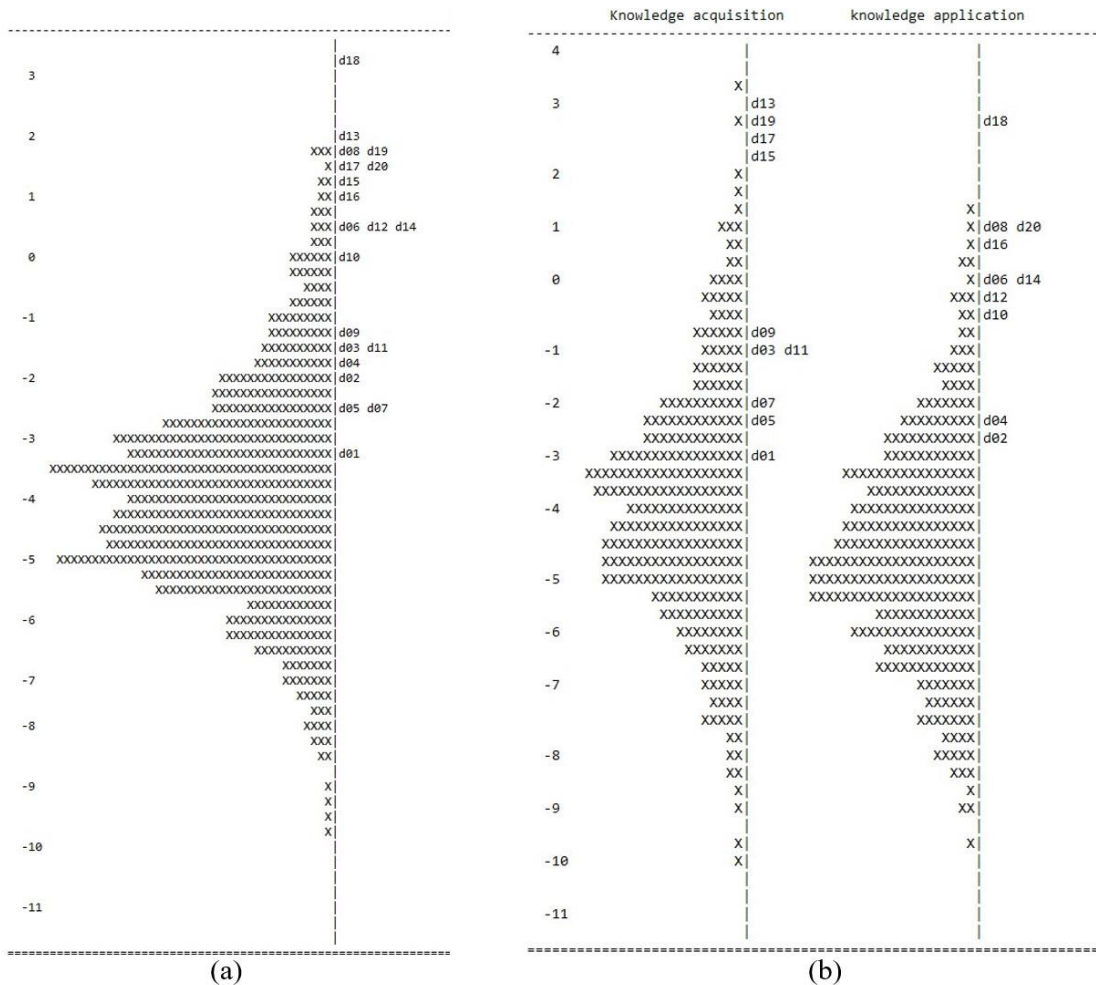


Figure 17. The Wright map of the MicroDYN test; (a) unidimensional (each X represents 2.1 cases); (b) two-dimensional (each X represents 4.4 cases)

Different item functioning (DIF) analysis was added to confirm the potential of test bias from the gender group. All items in MicroDYN tests showed no significant result for DIF ( $p > .05$ ), confirming that DIF did not occur concerning gender. The significant result is supported by the meaningful effect size, which DIF contrast not exceeds 0.64, except item 19 (Figure 18). However, due to nonsignificant result, item 19 is also considered to be independent from gender bias.

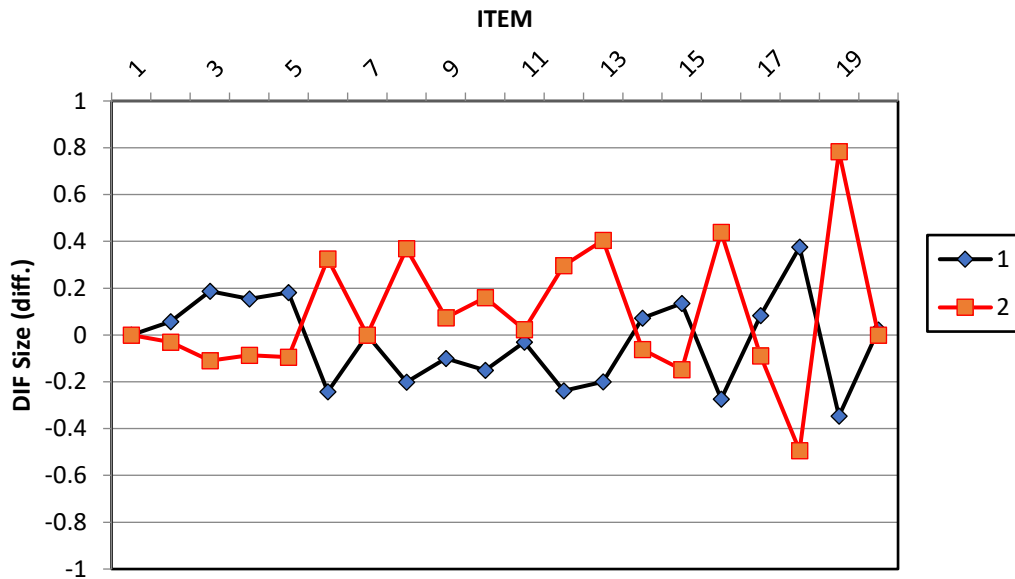


Figure 18. The DIF results for MicroDYN test (1 = male; 2 = female)

### 5.1.2 The ability level of the student’s domain-general problem-solving

We used the logit value to describe the general problem-solving ability. The overall student performance has a mean logit value of -3.93 (SD = 1.80) which shows a lower level of ability than the average (below logit 0) (see Figure 19). In knowledge application, the mean logit value of the students reaches -3.77 (SD = 2.18). This result also applies to knowledge application when the average student score is -3.98 (SD = 1.46). Additionally, we analysed the ability of the students based on the complexity of the test. In a simple task, the students got a mean logit score of -1.85 as their score dropped to -3.78 logit in a complex task and reached -3.40 logit for eigendynamic tasks. A similar trend also appears in the knowledge acquisition and knowledge application tasks, where the students' scores decreased steadily as the task became more complex and difficult (Figure 20).

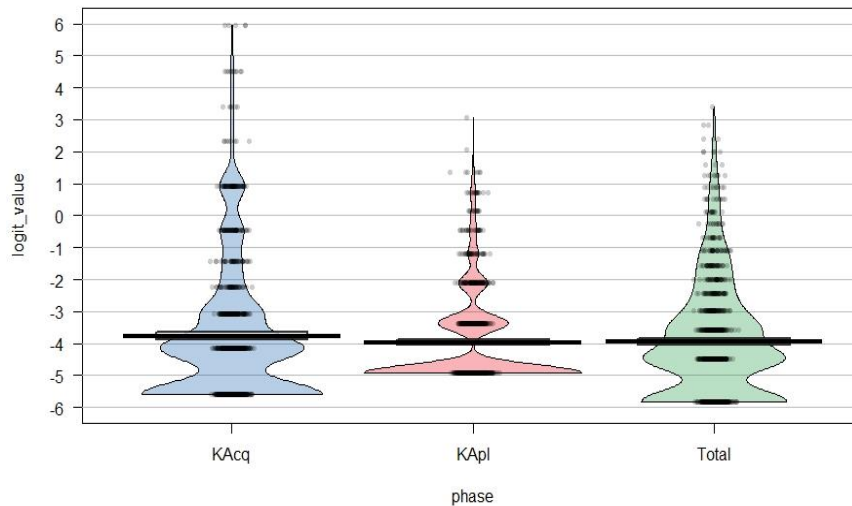


Figure 19. The distribution of students' logit scores in the MicroDYN test (Kacq = knowledge acquisition, Kapl = knowledge application)

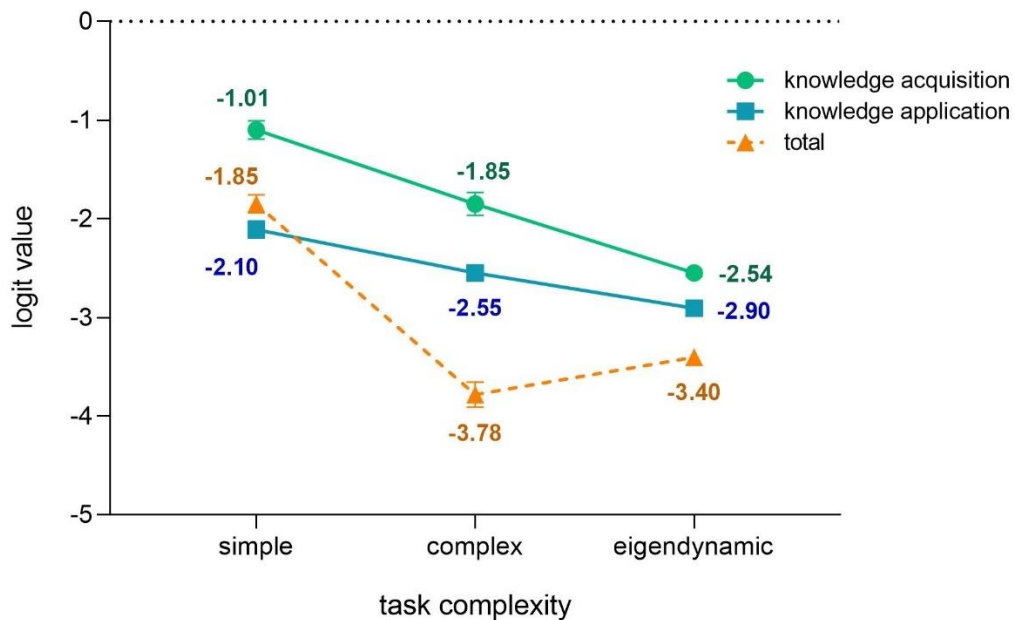


Figure 20. The distribution of students' problem-solving scores in tasks of different complexity

### 5.1.3 The gender difference in domain-general problem-solving ability

We elaborate on the distribution of students' abilities by analysing their gender differences in problem-solving tasks. Both male and female students exhibit low domain-general problem-solving ability with a mean logit value below 0 (see Figure 21). The detailed comparison between male and female students' problem-solving ability is presented in Table 12. The t-test analysis revealed a significant difference between male and female performance in problem solving, where male students performed better than female students ( $M_{\text{male}} = -3.58$ ,

$SD_{\text{male}} = 2.09$ ;  $M_{\text{female}} = -4.13$ ,  $SD_{\text{female}} = 1.58$ ;  $t = 5.18$ ,  $p < .001$ ). The male students also found to apply knowledge acquisition in better ways than the female students ( $M_{\text{Kac-male}} = -3.58$ ,  $SD_{\text{Kac-male}} = 2.09$ ;  $M_{\text{Kac-female}} = -4.13$ ,  $SD_{\text{Kac-female}} = 1.58$ ;  $t = 5.18$ ,  $p < .001$ ). In the knowledge application task, male students also scored better than female students ( $M_{\text{Kap-male}} = -3.66$ ,  $SD_{\text{Kap-male}} = 1.74$ ;  $M_{\text{Kap-female}} = -4.16$ ,  $SD_{\text{Kap-female}} = 1.22$ ;  $t = 5.18$ ,  $p < .001$ ). The favour of males compared to females is also found in different test complexity. In the simple MicroDYN task, male students scored better than female students, whereas it also applies for complex and eigendynamic tasks.

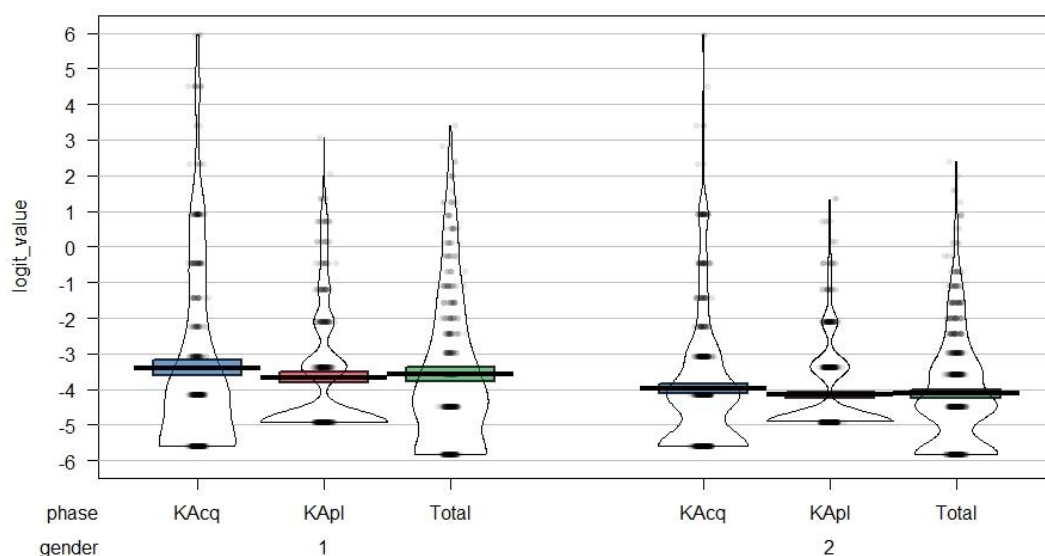


Figure 21. The distribution of students' domain-general problem-solving (Kacq = knowledge acquisition, Kapl = knowledge application) by gender (1 = male; 2 = female)

Table 12. The gender comparison of students' domain-general problem-solving

Category	Male (n = 457)			Female (n = 786)			t-test (df = 1241)	Cohens' d
	M <sub>score</sub>	M <sub>logit</sub>	SD	M <sub>score</sub>	M <sub>logit</sub>	SD		
MicroDYN	0.15	-3.58	2.09	0.09	-4.13	1.58	t = 5.18, p < .001**	0.38
Knowledge acquisition								
Simple	0.32	-0.89	1.80	0.26	-1.22	1.55	t = 3.34, p < .001**	0.18
Complex	0.26	-1.52	2.34	0.17	-2.04	1.88	t = 4.31, p < .001**	0.25
Eigendynamic	0.04	-2.46	0.83	0.01	-2.60	0.44	t = 3.93, p < .001**	0.24
Total	0.19	-3.40	2.56	0.13	-3.98	1.89	t = 4.61, p < .001**	0.29
Knowledge application								
Simple	0.20	-1.85	1.79	0.14	-2.26	1.36	t = 4.59, p < .001**	0.23
Complex	0.08	-2.35	1.25	0.03	-2.66	0.62	t = 5.97, p < .001**	0.29
Eigendynamic	0.04	-2.79	0.77	0.02	-2.97	0.44	t = 5.23, p < .001**	0.20
Total	0.10	-3.66	1.74	0.05	-4.16	1.22	t = 5.93, p < .001**	0.36

#### 5.1.4 Discussion

The MicroDYN test for domain-general problem-solving analysis has been proven to show good validity results in many context studies (Schweizer et al., 2013; Wu & Molnár, 2021; Wüstenberg et al., 2014). In the present study, the Indonesian version of the MicroDYN test shows acceptable psychometric properties for the validity of the construct and the internal consistency, which confirmed the applicability of the test for assessment purposes.

The assessment of domain-general problem-solving with Indonesian students resulted in an average score below logit 0 ( $M = -3.93$ ;  $SD = 1.80$ ), indicating poor performance in domain-general problem-solving tasks. A similar pattern is also found in both phases of problem solving (knowledge acquisition and knowledge application) when the students have an average score below logit 0. This result implies that students have difficulty in acquiring and applying knowledge in problem-solving tasks. During system exploration, most of the students failed to get new knowledge, find the relationship between input and output variables, and apply the knowledge in new situations. Consequently, their scores are dropping along with the higher complexity level of the task. The students get scores below 0 logit in the simple tasks and achieve lower in the complex and eigendynamic tasks.

Regarding the gender comparison, we found male students have better performance in domain-general problem-solving rather than female students. This result supports the hypothesis (H1) that boys outperformed girls in problem solving (Greiff et al., 2018; OECD, 2014; Wüstenberg et al., 2014). Borgonovi & Greiff (2020) analysed the PISA data and concluded that male students performed significantly better than female students in 19 out of 42 countries. The explanation for the differences among gender in problem solving lies in the effectiveness of using strategies during the knowledge-building process when facing unknown problem (Wittmann & Hattrup, 2004). Male students use more effective strategies in dealing with problem-solving tasks. They were able to analyse the role of each variable and altered those variables to reach a correct relation of input and output variables, making a clear system structure. In the cognitive psychology approach, male students have been found to show better spatial and information processing than female students (Reilly & Neumann, 2013) which supports their effectiveness during problem-solving task completion. Male and female students also have different styles in organising experiences and interpreting situations (Witkin & Goodenough, 1981) which influence their interaction with problem-solving properties and variables.

Since there is a lack of information on student general domain problem-solving performance in Indonesia (Wicaksono & Korom, 2022b), the present study will provide beneficial information for educational evaluation purposes which describe what students can do when dealing with general problem scenarios. This information can be used for further problem-solving programme in the scope of the assessment and intervention in high schools in Indonesia. Thus, we suggested to incorporate the acquisition and application of knowledge during learning activities. Later on, the studies of problem-solving assessment in the general domain become essential to prepare students to solve their real-life problems. The present study also has some limitations for using CPS approaches with dichotomous scoring, whereas the other domain-general problem-solving approaches with different scoring types are presentable. In terms of sample distribution, this study was dominated by female students. Indeed, for further study, the use of approximately equal number of sample groups is encouraged.

MicroDYN tests showed good psychometric properties for domain-general problem-solving assessment in the Indonesian context. Through the assessment process, we revealed the low level of student ability in domain-general problem solving. When performing the MicroDYN task, students' scores decrease along with the increasing complexity of the task. Therefore, we found that the result supports the hypothesis that male students outperformed female students in domain-general problem-solving tasks. Male students scored higher in both the knowledge acquisition and knowledge application categories. Thus, these results should be considered in terms of sample equity and the use of MicroDYN with CPS approaches for examining students' domain-general problem-solving, whereas other instruments may be applicable also for assessment purposes. Furthermore, we recommend further research on domain-general problem-solving in the area of assessment and intervention in the Indonesian context, by incorporating problem-solving phases and considering several factors that potentially influence problem-solving.

## 5.2 The assessment of domain-specific problem-solving

The present study aimed to measure students' domain-specific problem-solving which specifically addressed the science context. We used science problem-solving test to analyze students' abilities through the entire assessment process. Firstly, we check the validation of the science problem-solving test to ensure that the test fits the Indonesian sample. Secondly, we examined the students' abilities to solve science problems and made a comparison based on gender differences. The second hypothesis (H2) was tested in this subchapter, predicting that there is a significant difference between genders in domain-specific problem-solving performance, with males outperforming females (Gok, 2014; Soto-Ardila et al., 2022).

### 5.2.1 The validation of science problem-solving test

The science problem-solving test showed acceptable reliability (Cronbach's  $\alpha = .726$  and McDonald's  $\omega = .734$ ) confirming that the test has good internal consistency. We conducted CFA with a two-factor analysis that resulted in a good fit of the model with  $\chi^2/df = 1.99$ , CFI = .98, TLI = .98, RMSEA = .03, and SRMR = .04. The factor loadings for each item are above 0.40, except for item 1 (0.28) and item 3 (0.37) which belong to IP category. However, this item is still taken to explain IP based on the overall fit of the model. Thus, the Rasch analysis was conducted to support the position of those items in IP category, as well as the other items in each group. We performed unidimensional and multidimensional Rasch analysis to check the individual item fit. Table 13 shows the result of the Rasch analysis, confirming that each item in the science problem-solving test has a good fit index in the unidimensional model (infit MNSQ between 0.81–1.24). Furthermore, based on multidimensional model analysis, the infit MNSQ of the item also showed an acceptable value in the IP (0.86–1.11) and GS (0.93–1.09), categories claiming that each item fit each category. Another indicator of good item fit is the positive value shown in the point-biserial correlation (PTMA).

The Wright map in Figure 22 shows the distribution of students ability by 'X' (left) and the level of items (right). Among the eleven items in science problem-solving test, item 03 serve as the hardest items (logit measure = 0.92), while the easiest one is item 08 (logit measure = -0.94) which both of them belong to identifying problem category. Furthermore, differential item functioning (DIF) was performed to check the bias of the item toward the gender group that resulted in a nonsignificant result for the majority of the item, except item 7, 8, 9, and 10 ( $p < .05$ ). Further identification for item bias requires the estimation of DIF size indicating negligible result for bias if the value is higher than 0.64 (Boone et al., 2014). Thus, all of the items in the science problem-solving test have DIF size lower than 0.64 which also claims that

all of the items are free from bias, even for items 7, 8, 9, and 10 which is considered negligible (see Figure 23).

Table 13. The item fit of science problem-solving test

Variable	Items	Total score	Unidimension			2-dimension			PTMA
			Measure	SE	Infit MNSQ	Measure	SE	Infit MNSQ	
Identifying problem (IP)	d01	277	0.81	0.08	1.24	0.64	0.05	1.11	0.32
	d03	259	0.92	0.08	1.13	0.74	0.05	1.06	0.37
	d05	388	0.19	0.07	1.08	0.09	0.05	1.03	0.46
	d08	627	-0.94	0.07	0.93	-0.91	0.04	0.93	0.58
	d09	539	-0.53	0.07	0.81	-0.55	0.10	0.86	0.64
Generating solution (GS)	d02	471	-0.22	0.07	1.09	-0.13	0.05	1.09	0.47
	d04	497	-0.34	0.07	0.94	-0.24	0.05	0.99	0.56
	d06	316	0.58	0.08	1.01	0.61	0.05	1.01	0.48
	d07	379	0.24	0.07	1.05	0.29	0.05	1.06	0.46
	d10	530	-0.49	0.07	0.89	-0.39	0.05	0.93	0.60
	d11	474	-0.23	0.07	0.88	-0.14	0.10	0.93	0.59

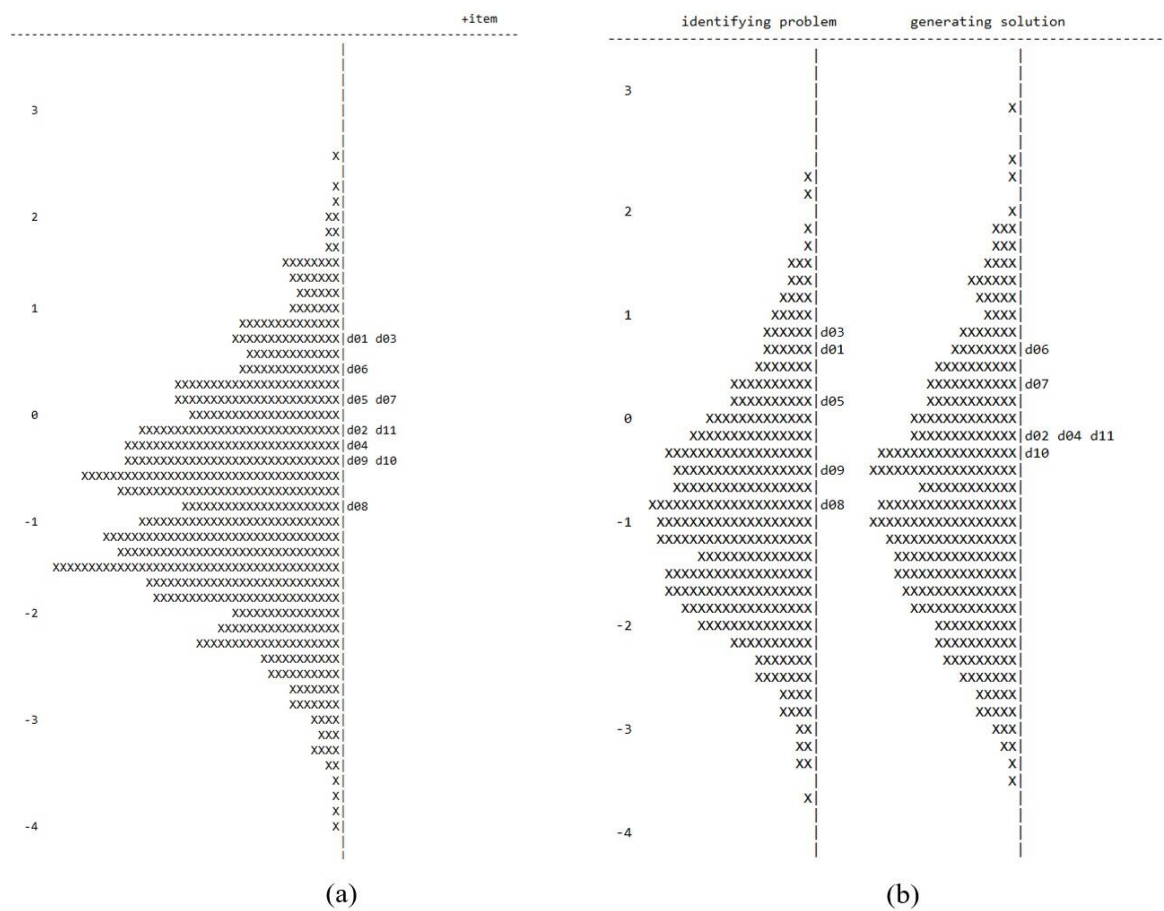


Figure 22. The Wright map of the science problem-solving test; (a) unidimensional (each X represents 2.0 cases); (b) multidimensional (each X represents 3.9 cases)

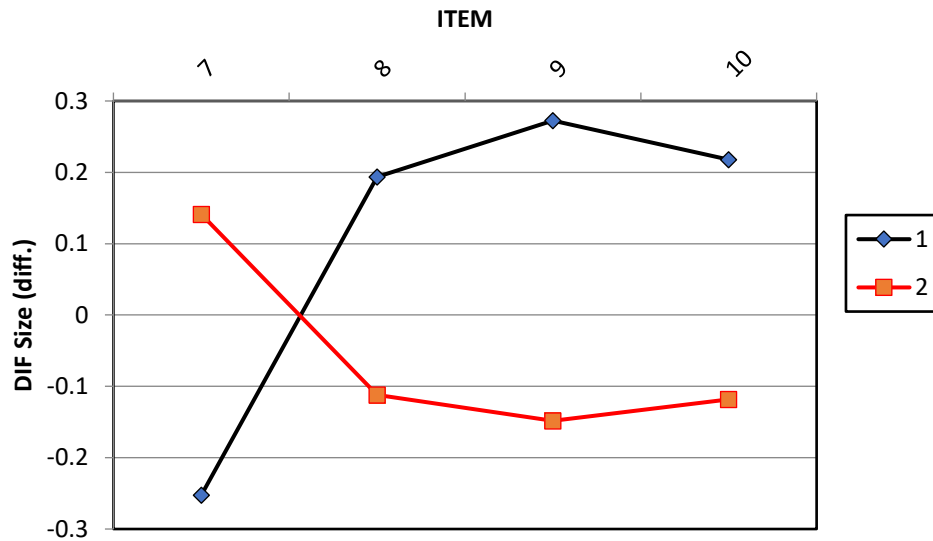


Figure 23. The DIF result for science problem-solving test items (1 = male; 2 = female)

### 5.2.2 The ability level of students' domain-specific problem-solving

The profile of the ability of the students to solve science problems is seen in Figure 24. The mean average of the students' scores is  $-0.88$  logit (SD = 1.43). In IP category, the students received average logit values of  $-0.93$  (SD = 1.51), which is lower than the average score of the GS category which reached  $-0.77$  logit (SD = 1.57). The distribution of the student score in SPS has a skewness value between 0.03 and 0.39 and a kurtosis value between  $-0.26$  and  $-0.62$ . Field (2013) reported that the normal distribution of the data can be accounted for with the skewness level within the range of  $\pm 1.96$ . Therefore, the SPS of the students in this study is normally distributed.

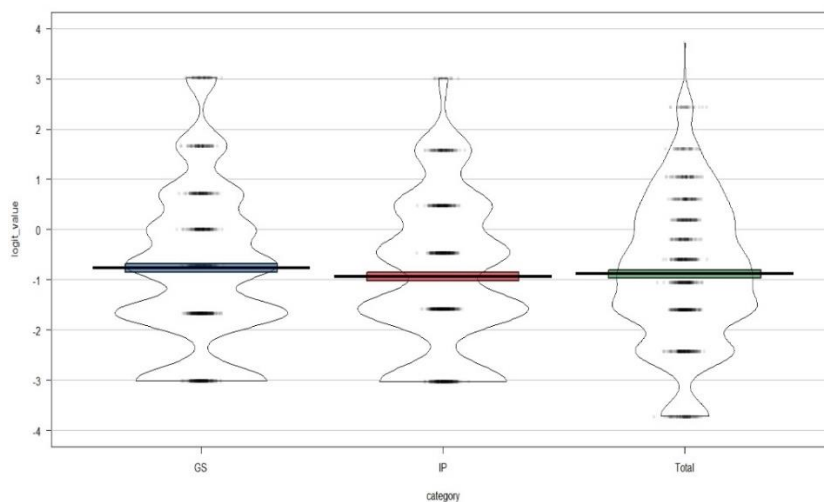


Figure 24. The distribution of students' logit scores in science problem solving (IP = identifying problem, GS = generating solution)

### 5.2.3 The gender difference in domain-specific problem-solving ability

The comparison of male and female students in the SPS score is presented in Figure 25. In SPS, female students ( $M_{\text{logit}} = -0.84$ ,  $SD = 1.39$ ) showed a slightly higher logit score than male students ( $M_{\text{logit}} = -0.96$ ,  $SD = 1.49$ ), but the difference is not significant ( $t = -1.36$ ,  $p > .01$ ) (see Table 14). We also compared the ability of male and female students in the SPS category, which resulted in no differences between male and female students in IP ( $t = -1.97$ ,  $p > .01$ ) and GS ( $t = -0.82$ ,  $p > .01$ ).

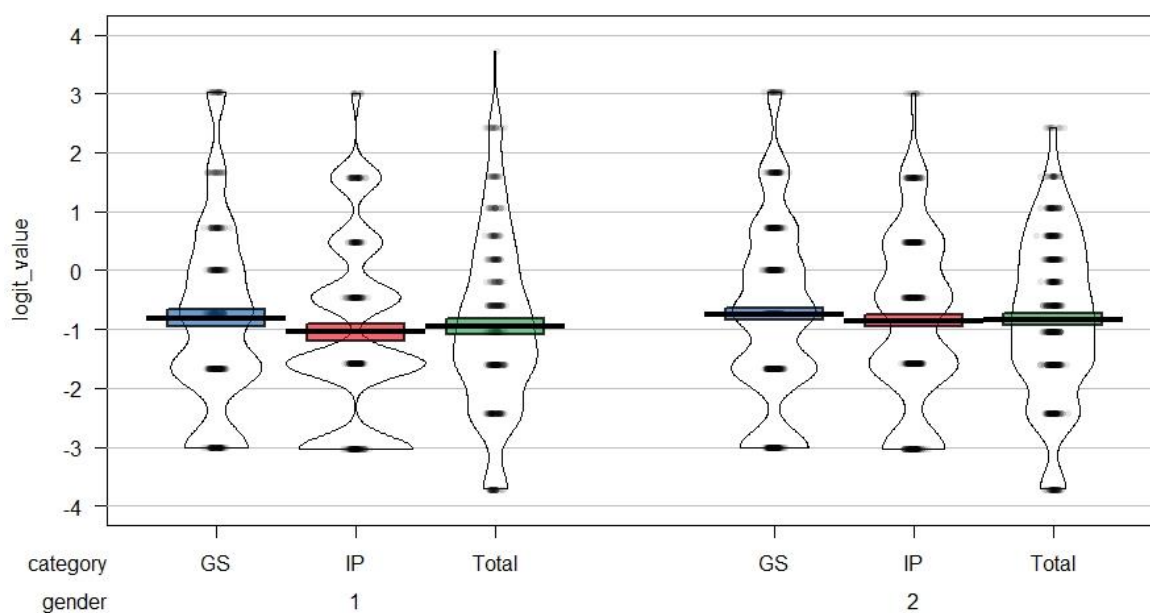


Figure 25. The distribution of students' science problem-solving ability (IP = identifying problem, GS = generating solution) by gender (1 = male; 2 = female).

Table 14. The gender comparison of students' science problem-solving

Category	Male (n = 457)			Female (n = 786)			t-test (df = 1241)	Cohens' d
	$M_{\text{score}}$	$M_{\text{logit}}$	SD	$M_{\text{score}}$	$M_{\text{logit}}$	SD		
IP	0.32	-1.04	1.57	0.35	-0.87	1.47	$t = -1.97, p = .054$	-0.11
GS	0.35	-0.82	1.59	0.36	-0.74	1.56	$t = -0.82, p = .413$	-0.07
Total	0.33	-0.96	1.49	0.35	-0.84	1.39	$t = -1.36, p = .166$	-0.08

Note: IP = identifying problem, GS = generating solution

### 5.2.4 Discussion

The dynamic of domain-specific problem-solving in adolescence has been studied by many scholars (Dermitzaki et al., 2009; Molnár et al., 2013; Rausch & Wuttke, 2016; Seifried et al., 2020; Walker et al., 2016). In the assessment process, the domain-specific problem-solving test

showed good psychometric properties in many contexts (Chang, 2010; Cheng et al., 2017; Wicaksono & Korom, 2023c). The present study also confirmed that SPS is suitable for domain-specific problem-solving assessment in science context in Indonesia. Based on the assessment result, students' domain-specific problem-solving is slightly below the average level with mean logit score as -0.88. In each domain-specific problem-solving task, students performed better in generating solution (GS) category rather than identifying problem (IP) tasks. It is also confirmed that the hardest item (item 3) belongs to the IP category explaining a problem situation in the agricultural topic. The IP task was designed to identify components or variables connected to the problem situation and to find its relationship for concluding the main cause of the problems. Furthermore, this category is connected to the reasoning process and the familiarity of the topics and prior knowledge. The low score in this category raises the probability that the majority of students have difficulty in reasoning activities and lack of essential knowledge.

The level of students' problem-solving ability in science reflects their habit in problem-solving activities. Priemer et al. (2020) explained that the success of problem solving in science involved a sequence of instructions or strategies during learning or a circular process of scientific activities. Thus, a wide spectrum of learning approaches is suggested to foster students' problem-solving activities (Pedaste et al., 2015), which can be used for teaching and learning process in Indonesia.

In case of gender difference, various results from previous studies have been presented regarding domain-specific problem-solving performance. Some studies showed that female students outperformed male students in domain-specific problem-solving, especially in mathematics and science contexts (Amalina & Vidákovich, 2023a; Ding et al., 2011). In contrast, other studies also revealed that male students have better scores than female students on mathematics, physics, and chemistry problem-solving tasks (Adigwe, 1992; Gok, 2014; Soto-Ardila et al., 2022). The present study revealed that there are no differences between the performance of male and female students in science problem-solving, rejecting the hypothesis (H2). This result, however, is supported by several studies that have reported no significant results in gender differences related to domain-specific problem-solving performance (Gallagher et al., 2000; Gunawan et al., 2020; Guven & Cabakcor, 2013; Harskamp et al., 2008; Hung et al., 2012). Regarding gender, many studies reported that male students develop more memory capacity to help them in problem representation and planning solution, while females are faster in verbal processing (Royer & Garofoli, 2005). Instead of the causing by cognitive processes, the domain-specific problem-solving of students is also related to learning

instruction and activities (Gok, 2014; Harskamp et al., 2008). In this study, the students expose the same learning activities, resulting in similar experiences in facing problem situations and performing solution plan. Although male and female students may exhibit different cognitive potential, through the learning activities, they tend to show similar performance in domain-specific problem-solving.

The result of study also implies that there is a need to improve students' problem-solving abilities, especially in the science area. Improvement can be suggested in the area of assessment and intervention, which incorporate problem-solving activities such as problem identification and solution-making during learning activities. It is also important to consider potential factors that have influence domain-specific problem-solving, such as affective and other cognitive variables. Furthermore, the result of this present study should be considered in light of limitations, where the proportion of the sample could be profound for gender equity. A further consideration is raised due to the use of science topics in the assessment of domain-specific problem-solving. The use of another specific topic for the assessment could potentially show yet a similar or different pattern in students' problem-solving performance. Thus, further study of domain-specific problem-solving within different contexts is encouraged. The present study exposes students' ability in domain-specific problem-solving in Indonesia.

We use a science problem-solving test to assess student domain-specific problem-solving that shows valid measurements with good psychometric properties. The assessment results in slightly below the average of the science problem-solving performance. We also found no significant difference between male and female students in domain-specific problem-solving, thus rejecting the hypothesis. The result raises a potential factor, such as similar experiences during science learning, which makes their ability levels much the same. However, this result should be considered with some limitation regarding the equality of gender size, the evaluation topic where the problem situation is about, and the specific context. Thus, further studies in domain-specific problem-solving assessment and intervention are encouraged with the involvement of various contexts and domains. Further recommendations also arise due to the participation of several variables that have potential to influence domain-specific problem-solving, including cognitive or affective factors.

### **5.3 The assessment of inductive reasoning**

This part of study aims to measure students' inductive reasoning ability. To ensure the validity of the assessment, we used confirmatory factor analysis and the Rasch model to verify psychometric properties of the test. Furthermore, we profiled the students' test score using a standardised logit value and compared the males' and females' score to obtain the complete distribution of the inductive reasoning ability. The third hypothesis (H3) was tested in this subchapter, predicting that there is a significant difference between gender in inductive reasoning performance (Blum et al., 2016), when males performed better mostly in figural and numerical tasks (Waschl & Burns, 2020)

#### **5.3.1 The validation of inductive reasoning test**

To ensure the validity of the inductive reasoning test, we performed a factor analysis and Rasch model to check the properties of the test. The reliability result of the test is good (Cronbach's  $\alpha = .895$  and McDonald's  $\omega = .894$ ) confirming the internal consistency of the test. The CFA with four factorial models also showed an acceptable fit value ( $\chi^2/df = 4.18$ , CFI = .95, TLI = .95, RMSEA = .05, and SRMR = .07) with a factor loading standardised between 0.53 and 0.98, except items 14 and 24 that showed a low loading value ( $< 0.40$ ). Further analysis with the Rasch model was added to check the quality of individual items. We performed unidimensional and multidimensional analysis for the test's validity and found that the test items showed a good fit for unidimensional (infit MNSQ = 0.83–1.30) and four-dimensional model (infit MNSQ = 0.85–1.26) (see Table 15). The fit of item is also indicated by the positive value in point-biserial correlation (PTMA). More information on the distribution of the items is presented in the Wright map (Figure 26). It showed that item NA8 acted as the most difficult item (logit value = 2.33) while item NS2 was the easiest item among all (logit value = -1.64).

Table 15. The item fit of the inductive reasoning test

Variable	Item	Total score	Unidimension			4-dimension			PTMA
			Measure	SE	Infit MNSQ	Measure	SE	Infit MNSQ	
Figure series (FS)	FS1	946	-0.86	0.08	1.05	-1.34	0.06	1.02	0.43
	FS2	812	-0.16	0.07	1.00	-0.58	0.05	0.92	0.49
	FS3	518	1.16	0.07	1.03	0.88	0.05	0.99	0.48
	FS4	723	0.25	0.07	0.95	-0.12	0.05	0.97	0.53
	FS5	453	1.45	0.07	1.11	1.21	0.05	1.15	0.40
	FS6	437	1.53	0.07	1.13	1.29	0.05	1.13	0.40
	FS7	731	0.21	0.07	0.90	-0.16	0.05	0.88	0.56
	FS8	920	-0.71	0.07	0.91	-1.18	0.14	0.85	0.53
Figure analogy (FA)	FA1	961	-0.94	0.08	0.92	-0.98	0.05	0.92	0.50
	FA2	910	-0.66	0.07	1.00	-0.69	0.05	0.96	0.46
	FA3	760	0.08	0.07	1.09	0.04	0.05	1.00	0.44
	FA4	799	-0.10	0.07	1.07	-0.14	0.05	0.99	0.45
	FA5	678	0.45	0.07	1.09	0.40	0.05	1.03	0.45
	FA6	588	0.85	0.07	1.24	0.80	0.05	1.18	0.34
	FA7	712	0.30	0.07	1.05	0.25	0.05	0.96	0.47
	FA8	696	0.37	0.07	0.86	0.33	0.13	0.87	0.59
Number analogy (NA)	NA1	788	-0.05	0.07	0.96	-0.77	0.05	1.02	0.50
	NA2	909	-0.65	0.07	0.87	-1.38	0.05	0.92	0.54
	NA3	551	1.01	0.07	1.14	0.31	0.05	0.99	0.41
	NA4	712	0.30	0.07	1.06	-0.41	0.05	0.93	0.47
	NA5	657	0.54	0.07	1.05	-0.16	0.05	0.92	0.47
	NA6	574	0.91	0.07	0.97	0.21	0.05	0.95	0.52
	NA7	500	1.24	0.07	0.96	0.55	0.05	0.93	0.52
	NA8	284	2.33	0.08	1.30	1.66	0.14	1.26	0.22
Number series (NS)	NS1	1059	-1.61	0.09	0.84	-0.71	0.07	0.98	0.49
	NS2	1063	-1.64	0.09	0.83	-0.75	0.07	1.00	0.50
	NS3	945	-0.85	0.08	0.88	0.36	0.07	1.00	0.53
	NS4	1005	-1.22	0.08	0.96	-0.16	0.07	1.13	0.46
	NS5	923	-0.73	0.07	0.91	0.54	0.07	1.02	0.52
	NS6	937	-0.81	0.08	0.84	0.43	0.07	0.94	0.55
	NS7	937	-0.81	0.08	0.83	0.42	0.07	0.92	0.56
	NS8	1000	-1.19	0.08	0.88	-0.12	0.18	1.01	0.51

As regards item bias, the majority of inductive reasoning items do not show significant results for DIF ( $p > .05$ ), except for some items (FS1, FS4, FS6, FS7, FA2, NA4, and NA5). For these items, we calculate the effect size to estimate whether the influence of gender is accounted as negligible or not.

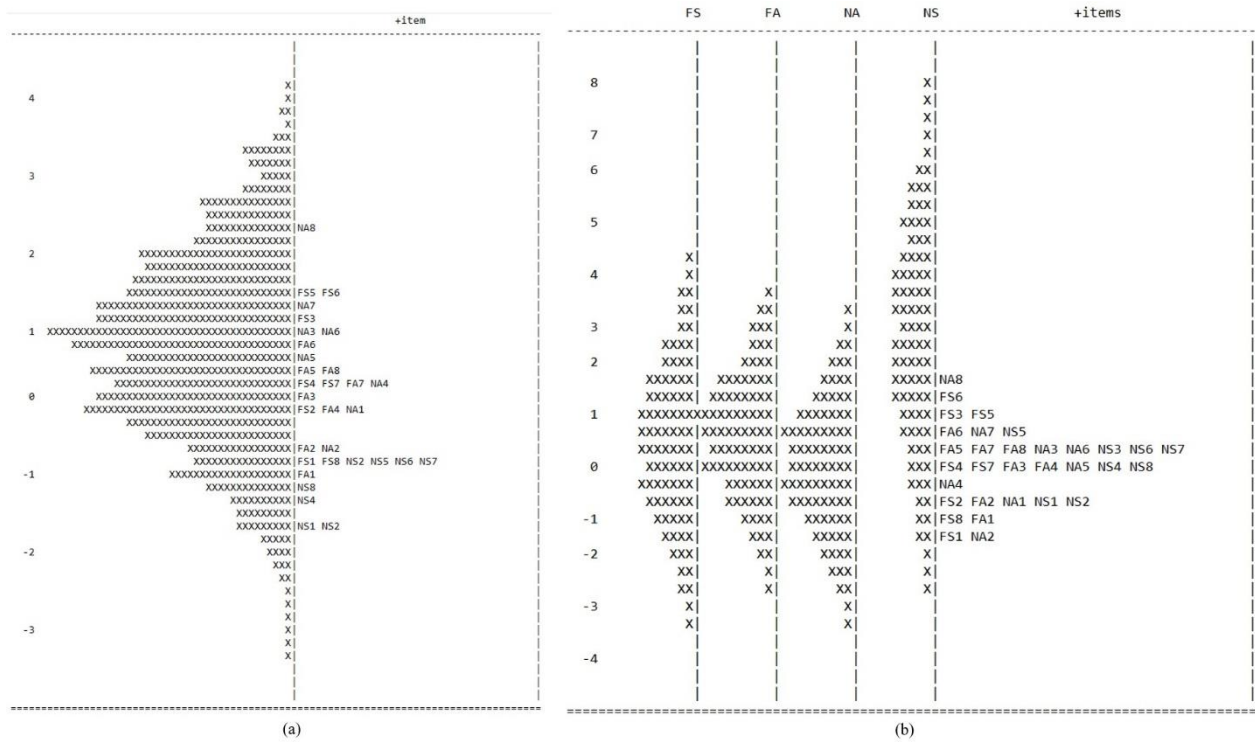


Figure 26. The Wright map of the inductive reasoning test; (a) unidimensional (each X represents 1.9 cases); (b) multidimensional (each X represents 14.2 cases)

All items showed a low DIF size ( $< 0.64$ ) indicating that the DIF is negligible for all test items (Figure 27). Indeed, we confirm that no potential bias is detected in the IR test.

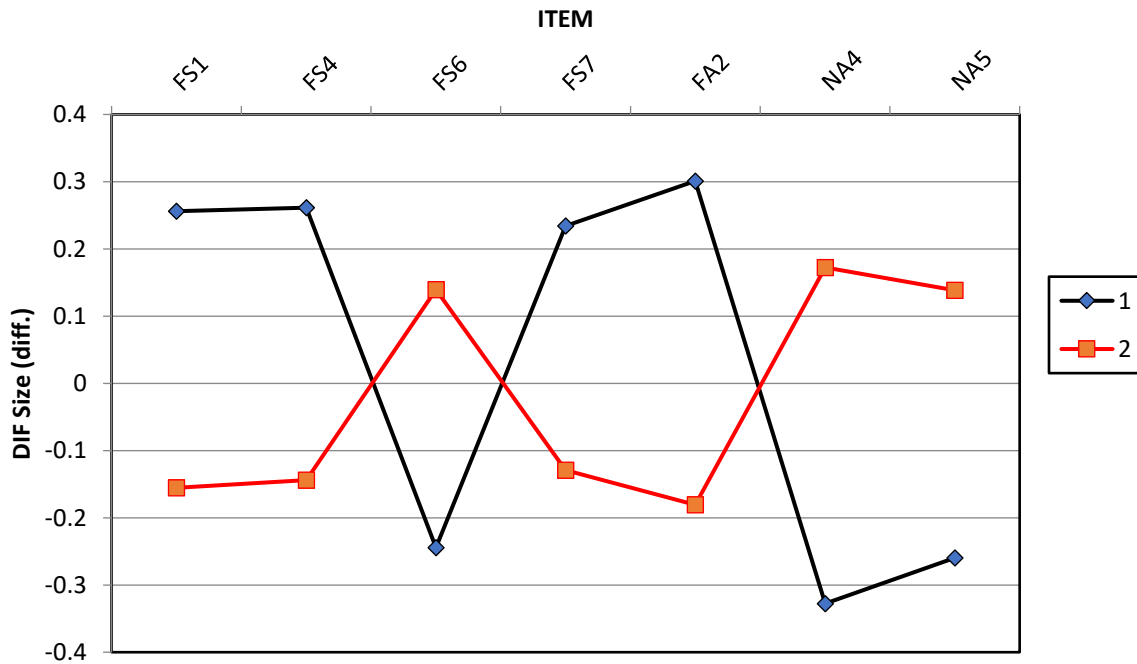


Figure 27. The DIF result for the inductive reasoning test (1 = male; 2 = female)

### 5.3.2 The ability level of student's inductive reasoning ability

The profile of the inductive reasoning ability of the students is seen in Figure 28. The mean average score of the students is 0.69 logit (SD = 1.41). In each category, students obtained a mean logit higher than 0 for figure series ( $M_{FS} = 0.36$ ;  $SD = 1.99$ ), figure analogy ( $M_{FA} = 0.69$ ;  $SD = 1.66$ ), and number series ( $M_{NS} = 1.89$ ;  $SD = 1.87$ ), while number analogy has a mean logit slightly lower than 0 ( $M_{NA} = -0.01$ ;  $SD = 1.76$ ). The distribution of the student score in the IR has a skewness value -0.06 and a kurtosis value of 0.56 which follow the criteria of normal distribution within the range of  $\pm 1.96$  (Field, 2013). Therefore, the students' inductive reasoning ability in this study is normally distributed.

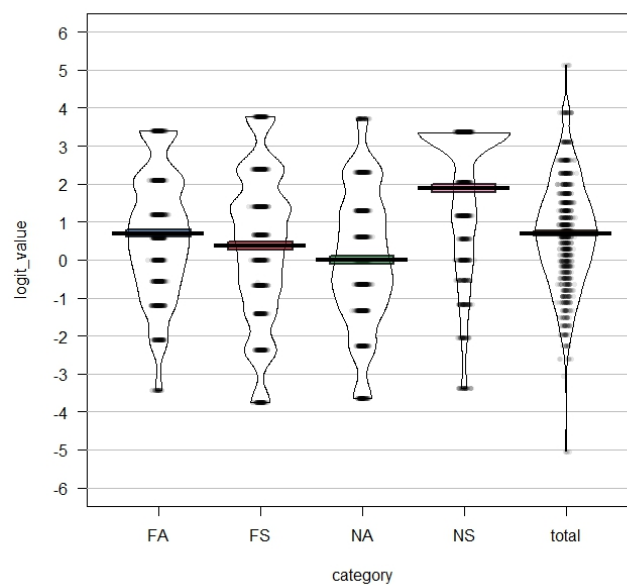


Figure 28. The distribution of students' logit scores in inductive reasoning (FA = figure analogy, FS = figure series, NA = number analogy, NS = number series)

### 5.3.3 The gender difference in inductive reasoning ability

We measure and compare the ability of male and female students in inductive reasoning. The distribution of students' inductive reasoning ability based on their gender is presented in Figure 29. The t-test comparison analysis results no significant difference between male and female students in inductive reasoning ( $M_{male} = 0.78$ ,  $SD_{male} = 1.54$ ;  $M_{female} = 0.65$ ,  $SD_{female} = 1.33$ ;  $t = 1.52$ ,  $p > .01$ ) (Table 16). In each category of tests, we found no differences between male and female students in the FS, FA, and NS ( $1.36 \leq t \leq 1.52$ ,  $p > .01$ ). Therefore, we found a significant difference between male and female students' ability in NA ( $t = 3.13$ ,  $p < .01$ ), when male ( $M_{male} = 0.20$ ,  $SD_{male} = 1.83$ ) has higher score than female ( $M_{female} = -0.13$ ,  $SD_{female} = 1.76$ ).

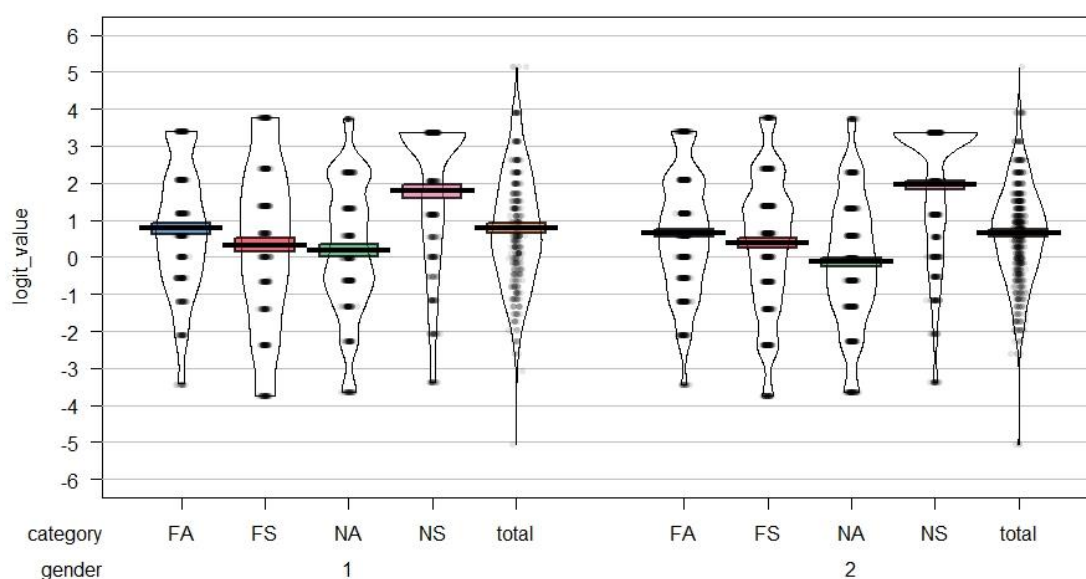


Figure 29. The distribution of students' inductive reasoning ability (FA = figure analogy, FS = figure series, NA = number analogy, NS = number series) by gender (1 = male; 2 = female)

Table 16. The gender comparison of students' inductive reasoning ability

Category	Male (n = 457)			Female (n = 786)			t-test (df = 1241)	Cohens' d
	M <sub>score</sub>	M <sub>logit</sub>	SD	M <sub>score</sub>	M <sub>logit</sub>	SD		
FS	0.55	0.32	2.19	0.61	0.38	1.86	t = -0.57, p = .571	-0.03
FA	0.63	0.77	1.73	0.61	0.64	1.62	t = 1.36, p = .174	0.04
NA	0.53	0.20	1.83	0.48	-0.13	1.76	t = 3.13, p = .002**	0.17
NS	0.78	1.79	1.95	0.80	1.95	1.82	t = -1.45, p = .147	-0.07
Total	0.62	0.78	1.54	0.61	0.65	1.33	t = 1.52, p = .129	0.04

Note: FA = figure analogy, FS = figure series, NA = number analogy, NS = number series \*\* significant at  $p < .01$

### 5.3.4 Discussion

This study presented the valid results of the inductive reasoning test for the high school level in the Indonesian sample. As in line with previous reports, the inductive reasoning test with figural and numerical tasks also showed good properties of validity and reliability in many studies (Molnár, 2011; Soeharto & Csapó, 2022; van de Vijver, 2002; Van Vo & Csapó, 2020; Wicaksono & Korom, 2022a). This means that the test can be used for a wider audience in different samples and contexts. Moreover, the use of figural and numerical tasks in the format of analogy and series completion was able to represent a reliable inductive reasoning measurement.

In comparison between inductive reasoning subtests, the students got the highest score in number series followed by figure analogy, figure series, and number analogy. The figural tasks comprise a collection of objects like circular and rectangular shapes that require students to recognise their pattern and size differences to find the relationship. Meanwhile, numerical tasks require mathematical recognition and computation to reveal relationship patterns (Liang et al., 2014). In this result, the students' tendency to perform better in a specific inductive reasoning task is unclear because both figural and numerical lean to the various patterns. However, while some studies have favoured figural tasks more than numerical tasks (Soeharto & Csapó, 2022; Van Vo & Csapó, 2020), this study captures another critical view in the format of the test, which includes series and analogy tasks. The students' performance is more varied when they solve series tasks rather than analogy tasks, which is confirmed by its higher standard deviation. Series tasks are aimed at ordering the attributes based on their complexity and patterns, while analogy represents information and objects information from previous experience into the new situation for making a judgement (Holyoak, 2012; Sternberg & Kalmar, 2004). In terms of analogy, students tend to show similar performance as a result of their previous learning experiences, whereas for series tasks, many students showed high performance, while some exhibited average and low performance. This may be related to students' individual cognitive development. Since both tasks require different cognitive processes, further information on students' cognitive ability is important to discover their dynamic in inductive reasoning performance.

As regards gender differences, there is no significant difference between male and female students in inductive reasoning, thus rejecting the hypothesis (H3). This result is supported by previous studies that have also reported nonsignificant results in gender comparison in inductive reasoning tasks (Díaz-Morales & Escribano, 2013; Kambeyo, 2018; Molnár, 2011; Van Vo & Csapó, 2021). Although in number analogy tasks, male students have a higher mean score compared to female students, their overall ability is at a similar level. Therefore, we found small effects favouring males toward females, and that the consistent measure of gender differences should be considered with variability in its magnitude and direction (Waschl & Burns, 2020).

This study provides the results of the inductive reasoning assessment in the Indonesian sample, suggesting that there are no gender differences in student performance. Although males showed a higher score in the number analogy task, the magnitude of gender size should account for further identification of inductive reasoning. The inductive reasoning tasks, involving both figural and numerical objects in analogy and series problem format, have been proven to be valid for the assessment process. However, some limitations should be considered due to the

availability of other tasks (verbal, matrices, scheme) which also play a significant role in describing students' inductive reasoning ability. Another concern is found in the distribution of tasks regarding their difficulty levels. As seen in Wright map, the number series tasks have a lower position compared to students' level distribution. Thus, it is recommended to add more difficult number series tasks in the item pool for better distribution of task difficulty.

The result of this study confirmed the validity of the measurement of inductive reasoning with figure series, figure analogy, number analogy, and number series tasks. These results indicated that the instrument can be used for further evaluation of inductive reasoning in a particular context. In Indonesia, most high school students performed better in inductive reasoning with an average score level above 0 logit. Among the tasks, students have higher diversity in series completion tasks rather than analogy tasks which is shown by its high standard deviation. The comparison analysis concerning gender showed no significant differences between male and female students, thus rejecting the hypothesis. However, a small effect still favours males, especially in number analogy tasks. Further research and investigation of inductive reasoning measurement is recommended with the involvement of other types of tasks such as verbal, matrices, and schemes. Additionally, for tasks in number series, additional items may be required to cover the distribution of difficult items. Thus, the identification of some influential factors may be necessary for further studies in inductive reasoning within the scope assessment or learning intervention.

#### **5.4 The assessment of scientific competency**

This part of the study aimed to investigate the scientific competency of Indonesian students. Firstly, the measurement test for scientific competency was designed based on the PISA science framework consisting of two main categories, explaining scientific phenomena and interpreting and evaluating scientific data or evidence. The Indonesian version consist of 15 items. Psychometric analysis was performed to validate the test. Furthermore, the distribution of the students' scientific competency was analysed based on standardised estimation and gender differences. The fourth hypothesis (H4) was tested in this subchapter, predicting that there is a significant gender difference in scientific competency, when females outperform males (OECD, 2004; Reilly, 2012).

### 5.4.1 The validation of scientific competency test

The scientific competency test showed acceptable reliability ( $\alpha = .775$  and  $\omega = .781$ ) confirming the internal consistency of the test between the participants. The construct validity analysis with the CFA results in a good fit of the model for the 2-factor scientific competency test ( $\chi^2/df = 1.99$ , CFI = .96, TLI = .96, RMSEA = .04, and SRMR = .05). The factor loading of each item is higher than 0.40 except for items SC08 (0.18), SC09 (0.24), and SC15 (0.33). A further analysis with the Rasch model was performed to support the fit of the individual item (Table 17).

Table 17. The item fit analysis results for scientific competency

Variable	Item	Total score	Unidimension			2-dimension			PTMA
			Measure	SE	Infit MNSQ	Measure	SE	Infit MNSQ	
Exploring and understanding phenomena (SCE)	SC01	598	-0.06	0.07	1.08	0.15	0.05	1.02	0.45
	SC02	598	-0.06	0.07	0.93	0.15	0.05	0.93	0.55
	SC04	804	-0.97	0.07	0.82	-0.69	0.05	0.87	0.60
	SC07	802	-0.96	0.07	0.94	-0.68	0.05	0.95	0.52
	SC09	465	0.53	0.07	1.28	0.69	0.05	1.19	0.31
	SC11	592	-0.03	0.07	0.88	0.17	0.05	0.89	0.58
	SC12	789	-0.90	0.07	0.98	-0.62	0.05	0.95	0.50
	SC15	431	0.69	0.07	1.16	0.83	0.12	1.09	0.36
Interpreting and evaluating scientific data (SCI)	SC03	655	-0.31	0.07	0.90	-0.53	0.05	0.92	0.57
	SC05	556	0.12	0.07	1.00	-0.12	0.05	1.02	0.50
	SC06	740	-0.68	0.07	0.92	-0.89	0.05	0.95	0.55
	SC08	143	2.47	0.10	1.16	2.11	0.05	1.14	0.21
	SC10	718	-0.58	0.07	0.87	-0.79	0.05	0.91	0.58
	SC13	414	0.77	0.07	1.00	0.49	0.05	1.03	0.48
	SC14	588	-0.03	0.07	1.07	-0.26	0.11	1.05	0.45

We perform unidimensional and multidimensional models to obtain comprehensive information about the fit of the item and support the result of CFA analysis. According to the unidimensional model, all items showed good fit indices (infit MNSQ = 0.82–1.28). In the 2-dimensional model, all elements also showed good fit indices in SCE (infit MNSQ = 0.87 – 1.19) or SCI (infit MNSQ = 0.91 – 1.14) with a positive value of the point biserial correlation (PTMA). A further distribution of the difficulty of the items was presented in the Wright map (Figure 30). The hardest item in the scientific competency test is SC08 (2.47 logit) while the easiest item is SC04 (-0.96 logit).

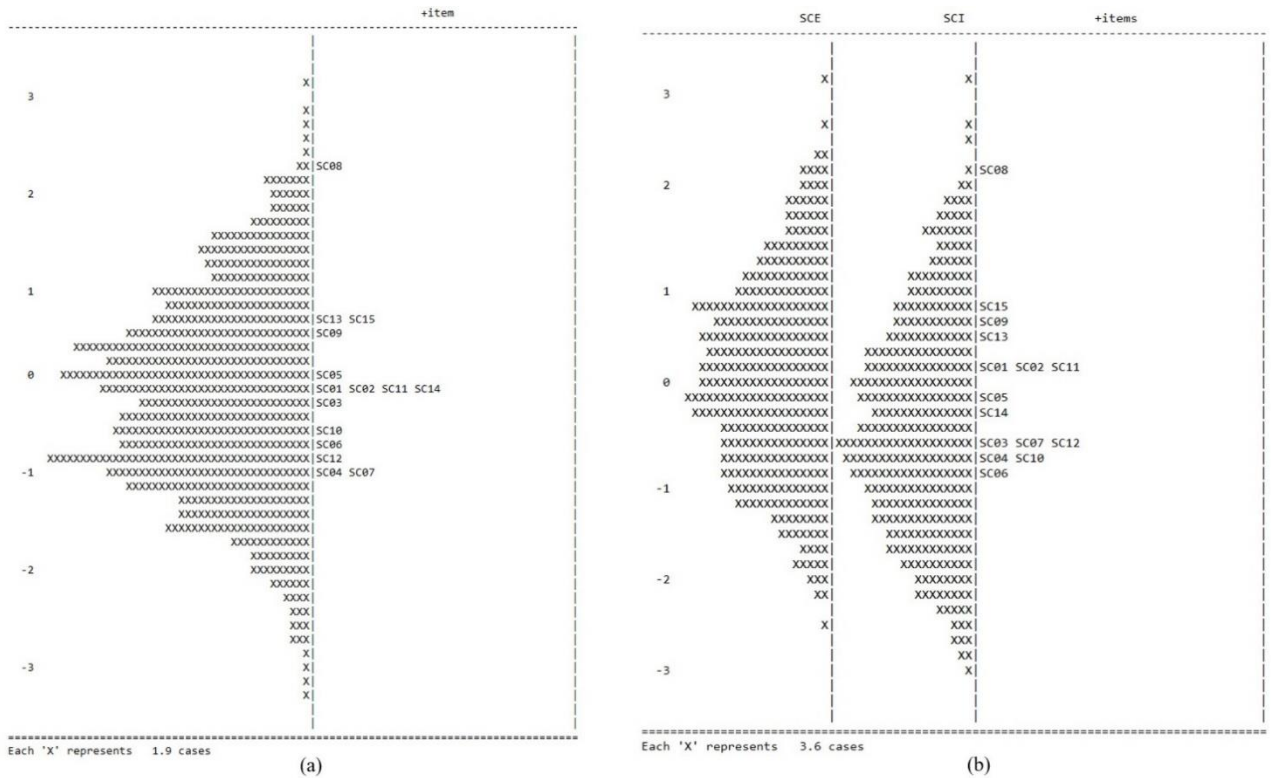


Figure 30. The Wright map of the scientific competency test items (a = unidimension, X represents 1.9 cases; b = multidimension, X represents 3.6 cases)

A further analysis of item bias was done with different item functioning (DIF). Bias was neglected if the DIF analysis showed a nonsignificant value or showed a high effect size  $> 0.64$ . Therefore, all items of the scientific competency test showed a nonsignificant result ( $p > .05$ ) with DIF size less than 0.64 (Figure 31), confirming that all items of the test have no bias towards the gender group.

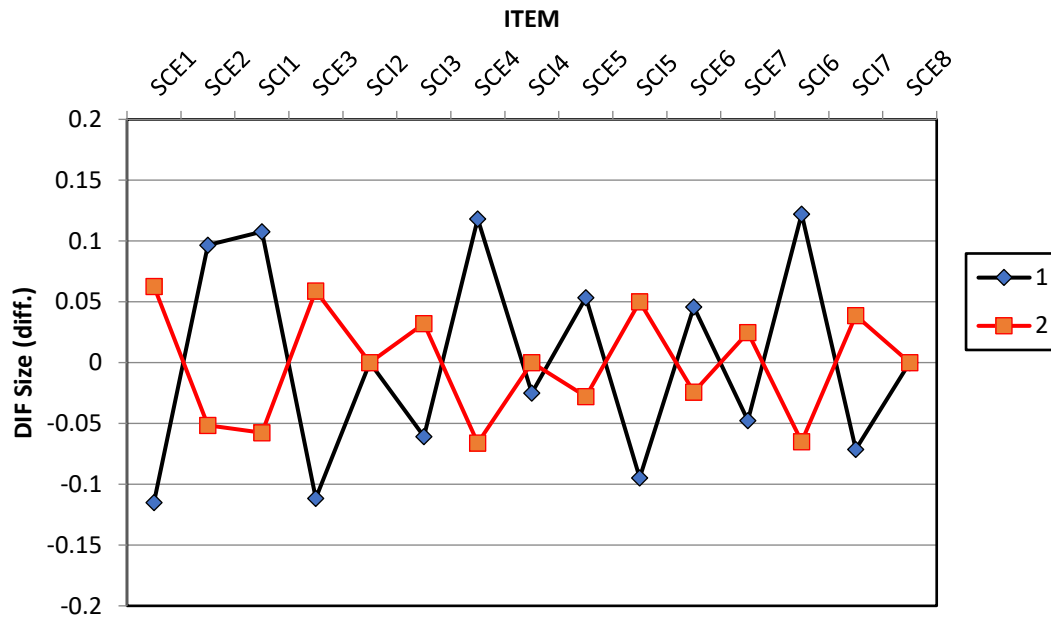


Figure 31. The DIF result of the scientific competency test (1 = male; 2 = female)

#### 5.4.2 The ability level of students' scientific competency

The student's ability to demonstrate scientific competency is shown in Figure 32. We used a standardised logit value from Rasch analysis to describe the level of students' ability. In the overall performance of scientific competency, Indonesian students reached a logit score close to 0 logit ( $M = -0.16$ ;  $SD = 1.34$ ), meaning that the ability is at an average level. In the SCE category, the students' score is slightly above 0 logit ( $M_{SCE} = 0.05$ ;  $SD = 1.48$ ), while the SCI category has a lower logit score ( $M_{SCI} = -0.43$ ;  $SD = 1.64$ ). The distribution of the students' score on the scientific competency test has a skewness value of 0.05 and a kurtosis value as -0.06 which follow the criteria of the normal distribution within the range of  $\pm 1.96$  (Field, 2013). Thus, the samples in this study are normally distributed.

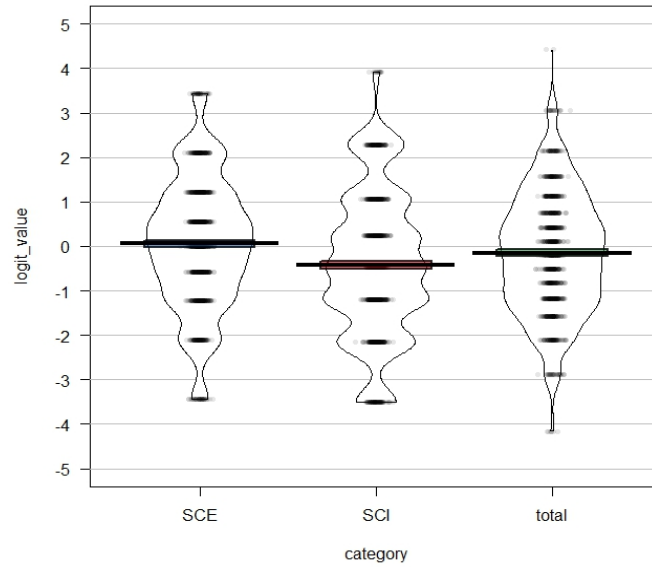


Figure 32. The distribution of students logit score in scientific competency (SCE = exploring and understanding; SCI = interpreting and evaluating)

### 5.4.3 The gender difference in scientific competency

We describe the scientific competency of the students between the male and female samples (see Figure 33), which results in a nonsignificant difference between the two groups (Table 18). In the overall SC test, male students received a mean logit of -0.21 (SD = 1.46) and -0.13 (SD = 1.26) for female students with no difference ( $t = -0.46, p > .05$ ). Furthermore, in each test category, we also did not find any significant differences between male and female students for SCE ( $t = -1.14, p > .05$ ) and SCI ( $t = -1.04, p > .05$ ).

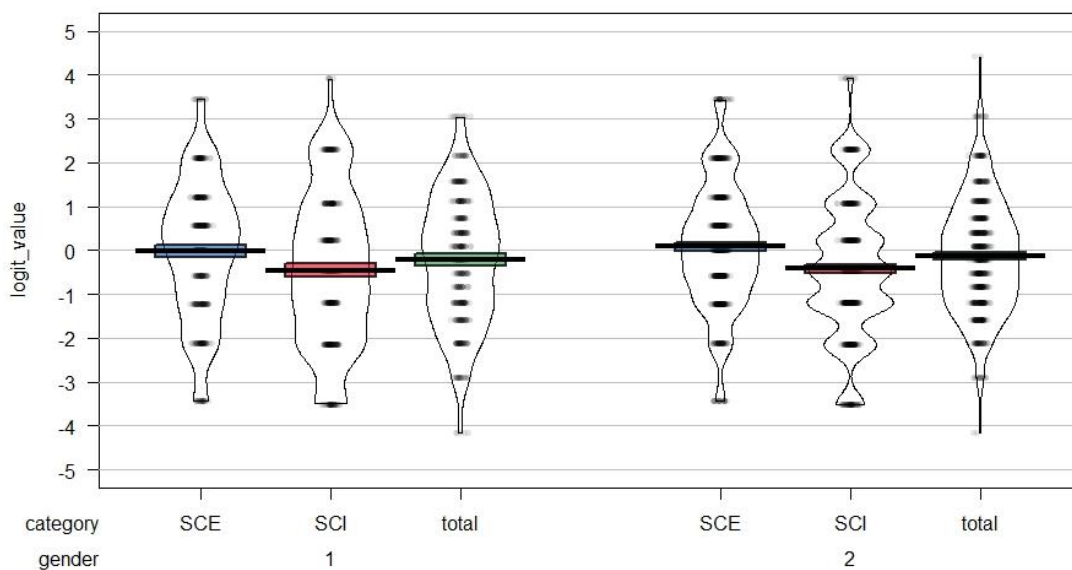


Figure 33. The distribution of students' scientific competency (SCE = exploring and understanding; SCI = interpreting and evaluating) by gender (1 = male; 2 = female)

Table 18. The gender comparison of students' scientific competency

Category	Male (n = 457)			Female (n = 786)			t-test (df = 1241)	Cohens' d
	M <sub>score</sub>	M <sub>logit</sub>	SD	M <sub>score</sub>	M <sub>logit</sub>	SD		
SCE	0.50	-0.01	1.57	0.51	0.09	1.42	t = -1.14, p = .257	-0.04
SCI	0.43	-0.46	1.74	0.44	-0.41	1.57	t = -1.04, p = .300	-0.04
Total	0.47	-0.21	1.46	0.48	-0.13	1.26	t = -0.46, p = .644	-0.04

Note: SCE = exploring and undeerstanding; SCI = interpreting and evaluating

#### 5.4.4 Discussion

The findings demonstrated that the psychometric analysis of the scientific competency test has good validity and internal consistency, which means that the test correctly measures the scientific competency of the students. The item test has a good distribution (see Figure 4.14) based on its difficulty level for both SCE and SCI categories. Since the test has been developed in response to the Indonesian content curriculum, it can be used for a further assessment in the Indonesian context, as well as for a global assessment in terms of its framework and content similarity. Furthermore, the scientific competency test with a focus on explaining scientific phenomena and interpreting and evaluating scientific data or evidence has been used in other studies and has been proven to yield a good validity result (Bybee et al., 2009; Lin et al., 2012; Tsai, 2015).

For the distribution of the students' performance in scientific competency, they achieved average level indicated by the position of logit value close to 0. From this result, it can be said that the students have a 50% chance of solving or failing the test. In each category of scientific competence, exploring and understanding phenomena (SCE) showed a higher value slightly above 0 logit compared to interpreting and evaluating scientific data (SCI) with a mean average below 0 logit. This result indicated that the majority of students are able to explore and understand scientific phenomena, but have more difficulties when they have to make interpretation or evaluation based on the given data or facts. The ability to explore and understand phenomena is categorised as low cognitive level. It involves information recalling and technical representation that include interpretation and extrapolation of knowledge or information (Anderson & Krathwohl, 2001; Köksal & Ulum, 2018). On the contrary, evaluation and interpretation require high cognitive demands because they require a judgement based on the criteria developed by the students themselves from the source information or knowledge they have acquired. When students make an evaluation, they pre-established criteria and made

distinction between conceptual and practises and draw a conclusion from their opinion (Granello, 2001).

In relation to gender differences, there is no significant difference between male and female students in scientific competency, as well as in each category, rejecting the hypothesis (H4). Although several studies on gender differences favor males towards females (Nwosu & Ibe, 2014; Reilly et al., 2019) on the contrary, other study has found that females performed better than male in some science performance (Japelj Pavešić, 2008; Shahzadi & Nasreen, 2020). Furthermore, the non-significant difference between gender in science performance also reported by PISA in many countries that participated in the science survey (OECD, 2004). Furthermore, gender stereotypes related to cognitive performance have been proven, with males being better in spatial and computational abilities, as well as specific tasks requiring analytical and high cognitive processes, while female tend to perform better in verbal related tasks, memorisation, and rote learning (Hyde, 2016). Correspond to the result of this study, Reilly et al. (2019) also found that male students have higher variability in science performance, as indicated by the high standard deviation compared to female students, indeed their difference is not significantly high. Thus, the gender differences in connection to scientific competency vary based on context and the level of gender equity influences (Cheng et al., 2021). The other possible explanation also arises from the culture and learning environment that has been shown to influence students' performance in science (Fang & Wei, 2010; Luu & Freeman, 2011; Sung & Hwang, 2013).

The result of this study provides valuable information on the students' performance in scientific competency which categorises the students' ability at an average level. Indeed, this study suggest further improvement of educational practices for scientific competency in Indonesia. Like any study, there are some limitations in the measurement, mainly because the context of the test was mostly related to biology and natural science. Thus, the use of a broader context addressing other science-related topics is also relevant to assess students' scientific competency. This study specifically addressed secondary school students, which is consistent with the criteria for global assessment and compulsory education in Indonesia. Indeed, the participation of different age groups is recommended for further studies to get a complete picture of students' scientific competency profile. Furthermore, is recommended to identify other factors that potentially influence scientific competency for further study, as well as the exploration in the area of learning and instruction to promote scientific competency.

The use of the scientific competency test with two main categories (explaining scientific phenomena and interpreting and evaluating scientific data or evidence) shows an acceptable

result for validity and reliability analysis. The assessment result confirmed that Indonesian students achieved moderate level of scientific competency. No gender differences were found, indicating that males performed similarly to females, thus rejecting the hypothesis. However, this result must be interpreted in light of some limitations regarding the context of measurement and sample equity. Further improvement in terms of assessment is recommended by addressing other topic in science for the test's construction. It is also suggested to conduct research in the area of learning and instruction to promote student scientific competency as well as the exploration of cognitive and affective factors that potentially influence students' scientific competency.

## **5.5 The assessment of attitudes towards science**

In this part of study, we focused on attitude towards science as one of the affective factors associated with problem solving. We evaluated the student's attitude towards science with the attitude towards science questionnaire (ATSQ). The construction of ATSQ was done through adaptation and reconstruction of the existing questionnaire (three-dimensional framework) to determine the learning context in Indonesia. Thus, the validity analysis of the ATSQ was performed followed by the analysis of the attitude profile of the students and the differences between the gender groups. The fifth hypothesis (H5) was tested in this subchapter, predicting that there is a significant difference between gender in attitude towards science, when females outperform males (Aini et al., 2019; Liu et al., 2010; Mihladiz et al., 2011).

### **5.5.1 The validation of attitude towards science questionnaire**

To ensure the validity of ATSQ, we check the internal consistency of the questionnaire, resulting in an acceptable reliability coefficient (Cronbach's  $\alpha = .834$  and McDonald's  $\omega = .827$ ). For the validity of the construct, we run CFA, resulting in an acceptable fit of the model for the four factors of ATSQ ( $\chi^2/df = 7.46$ , CFI = .94, TLI = .94, RMSEA = .07, and SRMR = .05) with factor loading for each item greater than 0.40. Further analysis was performed to confirm the quality of the item with the Rasch model (Table 19). We checked the fit indices of each item in the unidimensional and multidimensional models (4 factors). The ATSQ infit MNSQ showed an acceptable range (0.61–0.49) in the unidimensional model, except for item A04 which has an infit MNSQ slightly above the range (1.51). In the 4-dimensional model, most items also have a good fit MNSQ (0.65–1.46), except items A08 (1.51) and A25 (1.62), which have a value slightly above the threshold. All items have a positive point

biserial (PTMA) correlation. Thus, we consider those items to remain in the questionnaire since the value does not exceed too far from the threshold (Bond & Fox, 2013).

Table 19. The item fit analysis results for the ATSQ

Variable	Item	Total score	Unidimension			4-dimension			PTMA
			Measure	SE	Infit MNSQ	Measure	SE	Infit MNSQ	
Enjoyment (ENJ)	A01	4168	-0.31	0.04	0.69	-0.38	0.03	0.82	0.63
	A03	3487	0.53	0.03	0.69	0.85	0.03	0.84	0.55
	A05	4105	-0.23	0.04	0.79	-0.26	0.03	1.05	0.51
	A09	3371	0.67	0.03	0.76	1.05	0.03	0.98	0.56
	A12	4008	-0.11	0.04	0.75	-0.08	0.03	0.82	0.63
	A20	4217	-0.37	0.04	0.63	-0.47	0.03	0.65	0.64
	A21	4227	-0.38	0.04	0.61	-0.48	0.03	0.68	0.64
	A26	4084	-0.20	0.04	0.64	-0.22	0.08	0.68	0.61
Anxiety and difficulty (ANX)	A04	3254	0.80	0.03	1.51	0.07	0.03	1.16	0.03
	A06	3430	0.59	0.03	1.40	-0.24	0.03	1.32	0.16
	A13	3783	0.17	0.04	1.36	-0.87	0.03	1.08	0.01
	A16	3023	1.08	0.03	1.42	0.47	0.03	0.95	0.08
	A17	3120	0.96	0.03	1.52	0.30	0.03	0.89	0.02
	A18	3169	0.91	0.03	1.51	0.22	0.03	0.96	0.02
	A23	3259	0.80	0.03	1.49	0.06	0.07	0.98	0.02
Participation in science learning and activities (PAR)	A07	3447	0.57	0.03	1.18	0.87	0.03	1.46	0.59
	A08	3768	0.18	0.04	1.16	0.30	0.03	1.51	0.57
	A14	4200	-0.35	0.04	0.92	-0.48	0.03	1.15	0.59
	A15	3772	0.18	0.04	0.81	0.30	0.03	1.02	0.60
	A22	3777	0.17	0.04	0.60	0.29	0.03	0.72	0.63
	A24	3942	-0.03	0.04	0.68	-0.01	0.03	0.92	0.57
	A25	3825	0.11	0.04	1.19	0.20	0.03	1.62	0.51
	A27	4744	-1.03	0.04	0.89	-1.48	0.08	1.22	0.51
Value of science (SVAL)	A02	4631	-0.89	0.04	0.87	0.43	0.03	1.08	0.53
	A10	5101	-1.50	0.04	0.93	-0.48	0.03	0.93	0.47
	A11	4842	-1.16	0.04	0.99	0.03	0.03	1.08	0.43
	A19	4850	-1.17	0.04	0.88	0.02	0.05	0.96	0.42

Additionally, we distributed the level of items by their logit measures (see Figure 34). Based on the Wright map, item A16 is classified as the most difficult item among ATSQ items with a 1.08 logit measure, while the easiest item is A10 (-1.50 logit). Further analysis was performed with different item functioning (DIF) to detect the item bias based on the gender group. DIF analysis showed that some elements of the ATSQ have a significant result that potentially showed bias towards gender (item A02, A04, A07, A08, A13, A14, A16, A17, A18, A21, A23, A27). Thus, we checked the DIF size of these significant items to ensure that their bias is negligible. Figure 35 shows the DIF size of the significant items that showed low size (< 0.64),

meaning that DIF is negligible for these items. These results confirm that all items in the ATSQ are independent of gender bias.

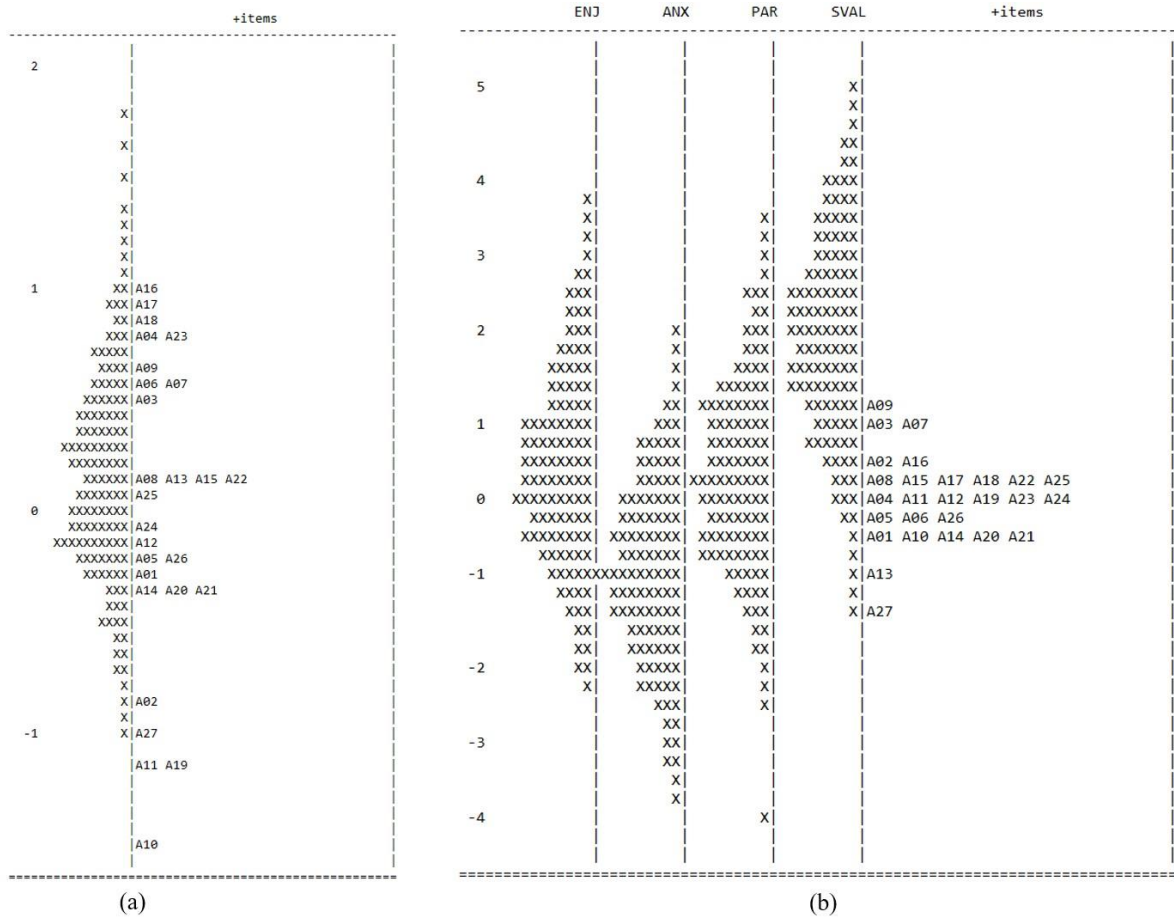


Figure 34. The Wright map of the ATSQ items (a = unidimension, X represents 2.0 cases; b = multidimension, X represents 10.8 cases)

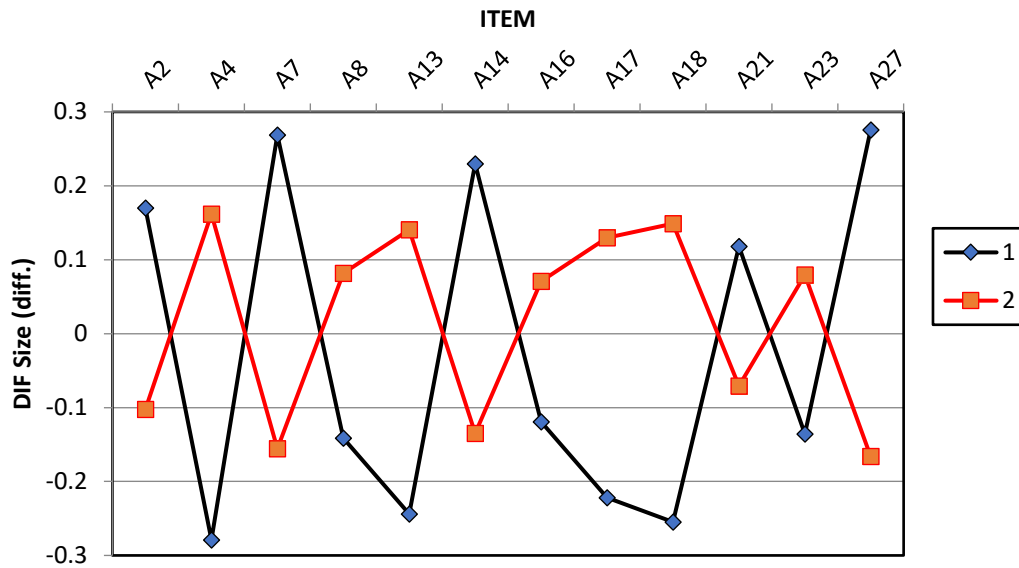


Figure 35. The DIF results for ATSQ (1 = male; 2 = female)

### 5.5.2 The profile of students' attitude towards science

We describe the students' attitudes towards science, which is illustrated in Figure 36. The overall students' attitude towards science is 0.21 logit (SD = 0.65) slightly above the average 0 logit. The highest score for attitude towards science was found in the SVAL variable (M = 2.36; SD = 2.05) implicating that students tend to give positive responses about the value of science in life and society. For the ENJ variable, the students' response is close to 0 logit (M = 0.49; SD = 2.15) as well as the PAR variable (M = 0.31; SD = 1.39). On the other hand, the ANX variable got the lowest score below logit 0 (M = -0.75; SD = 1.61) indicating the low level of anxiety and difficulty in science. The distribution of students' scores in attitudes towards science has a skewness value around 1.88 and a kurtosis value of 16.01 which does not follow the criteria of normal distribution. However, in each ATSQ category, the students' responses followed a normal distribution (skewness = -0.06–0.40; kurtosis = -0.41–2.18).

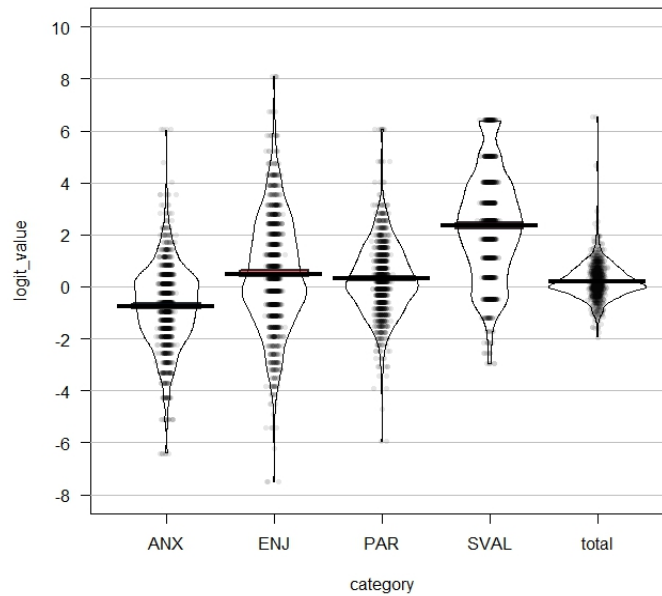


Figure 36. The distribution of students' logit scores in attitude towards science (ANX = anxiety and difficulty; ENJ = enjoyment; SVAL = value of science; PAR = participation in science activities)

### 5.5.3 The gender difference in attitude towards science

We compared the attitude towards science between male and female students (see Figure 37). The detailed result of the comparison analysis is presented in Table 4.20, which results in no significant difference in attitude towards science between genders ( $M_{\text{male}} = -0.19$ ,  $SD_{\text{male}} = 0.76$ ;  $M_{\text{female}} = 0.21$ ,  $SD_{\text{female}} = 0.58$ ;  $t = -0.66$ ,  $p > .05$ ). However, in each ATSQ variable, we found a significant difference between gender in ENJ with female scored higher than male ( $M_{\text{male}} = 0.33$ ,  $SD_{\text{male}} = 2.29$ ;  $M_{\text{female}} = 0.59$ ,  $SD_{\text{female}} = 2.06$ ;  $t = -2.05$ ,  $p < .05$ ). Female students also scored higher in PAR ( $M_{\text{male}} = 0.16$ ,  $SD_{\text{male}} = 1.46$ ;  $M_{\text{female}} = 0.40$ ,  $SD_{\text{female}} = 1.35$ ;  $t = -2.87$ ,  $p < .05$ ) and SVAL ( $M_{\text{male}} = 2.19$ ,  $SD_{\text{male}} = 2.16$ ;  $M_{\text{female}} = 2.46$ ,  $SD_{\text{female}} = 2.00$ ;  $t = -2.24$ ,  $p < .05$ ) compared to male students. On the contrary, for the ANX variable, male students have a higher score compared to female students ( $M_{\text{male}} = -0.49$ ,  $SD_{\text{male}} = 1.61$ ;  $M_{\text{female}} = 2.46$ ,  $SD_{\text{female}} = -0.91$ ;  $t = 4.52$ ,  $p < .05$ ), indicating that males are more anxious and face more difficulties than female in science.

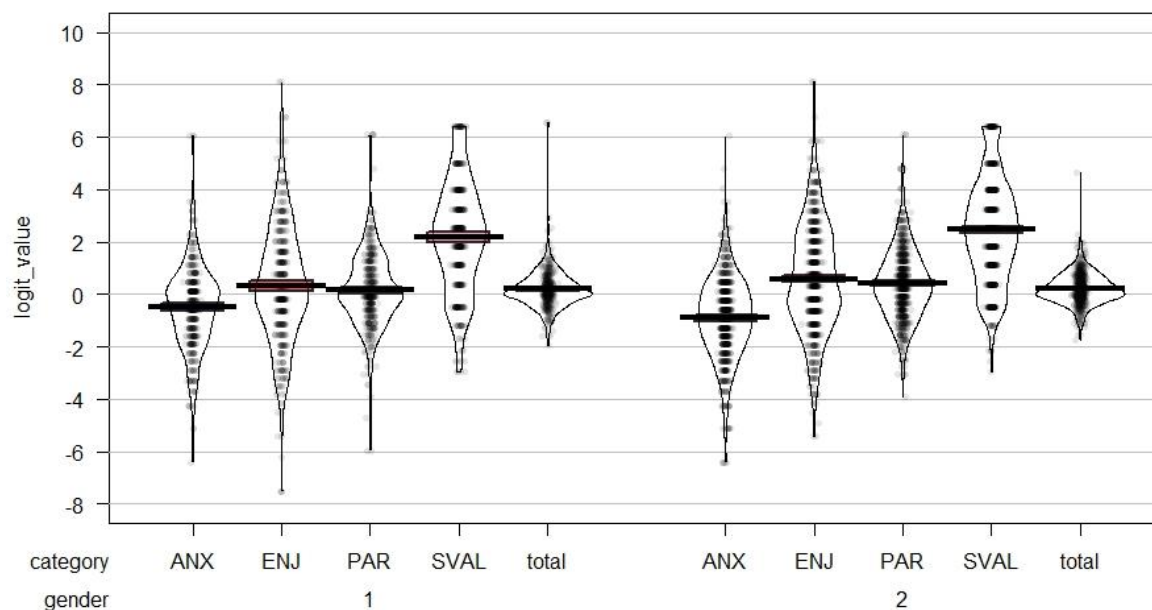


Figure 37. The distribution of students' attitude towards science (ANX = anxiety and difficulty; ENJ = enjoyment; SVAL = value of science; PAR = participation in science activities) by gender (1 = male; 2 = female)

Table 20. The gender comparison of students' attitude towards science

Category	Male (n = 457)			Female (n = 786)			t-test (df = 1241)	Cohens' d
	M <sub>score</sub>	M <sub>logit</sub>	SD	M <sub>score</sub>	M <sub>logit</sub>	SD		
ENJ	3.13	0.33	2.29	3.21	0.59	2.06	t = -2.05, p = .040*	-0.12
ANX	2.84	-0.49	1.61	2.71	-0.91	1.58	t = 4.52, p = .001**	0.24
PAR	3.09	0.16	1.46	3.21	0.40	1.35	t = -2.87, p = .004**	-0.19
SVAL	3.85	2.19	2.16	3.94	2.46	2.00	t = -2.24, p = .025*	-0.14
Total	3.23	0.19	0.76	3.27	0.21	0.58	t = -0.66, p = .512	-0.10

Note: ANX = anxiety and difficulty; ENJ = enjoyment; SVAL = value of science; PAR = participation in science activities. \*\* significant at  $p < .001$ , \*  $p < .05$

### 5.5.4 Discussion

The construct validity with CFA showed an acceptable model fit, indicating that the instrument has a good theoretical foundation with four latent variables. Rasch analysis is also considered one of the most powerful measurements to evaluate assessment tools with some functions in item positioning (Boone, 2016). It can determine the quality of the measurement by calculating the fit measure based on the level of the item and the responses of the person. The Indonesian version of ATSQ showed a prominent fit value for each item and acceptable reliability from the analysis of the Rasch model, indicating that the questionnaire measured the trait correctly and consistently. This result is also supported by some studies that confirm that

ATSQ with three-dimensional variables is reliable for attitude measurements (Wicaksono & Korom, 2023a, 2021). Furthermore, the DIF analysis showed that the measurement is not biased toward gender. Therefore, the ATSQ has overall good psychometric properties, which confirms that the questionnaire is reliable and valid to assess students' attitudes toward science.

This study confirms a moderate level of students' attitude towards science, as indicated by a mean score close to 0 logit. The higher attitude was found in the value of science category, confirming that students have strong beliefs about the relevance of science in daily life. This result also indicates that students understand the importance of science in society, which can be a potential factor in supporting their science-related studies. In the other categories, most of the students respond positively to enjoyment and participation in science learning and activities. A supportive environment such as a classroom and other learning activities is believed to be able to influence people's attitudes (Almasri et al., 2021; Krogh; & Thomsen, 2005). Furthermore, several studies also reported that the level of enjoyment in learning is connected to learning strategy and interactive activities (Chao et al., 2015; Ralph et al., 2022; Vandecandelaere et al., 2012). Indeed, if the students have a willingness to participate in science learning and activities, they tend to show enjoyment and a positive attitude towards science. Meanwhile, the anxiety and difficulty categories have the lowest score among all attitude variables, suggesting that students experience fewer negative feelings towards science and less difficulty with science in their current state. Anxiety and difficulty represent negative emotions that counterbalance other variables of attitude, especially enjoyment. Indeed, some studies also support that anxiety has a negative correlation with enjoyment, which makes students who have higher enjoyment tend to show less anxiety (Raccanello et al., 2019; Wicaksono & Korom, 2023a).

Regarding the gender comparison, there are no significant differences in the general attitude towards science, thus rejecting the hypothesis (H5). However, when it comes to enjoyment, participation in science learning and activities, and the value of science variables, female students showed a better score than male students. In the anxiety category, male students have a higher score compared to female students, indicating that female students tend to show a more positive attitude compared to male students (Khishfe & BouJaoude, 2016; Mihadiz et al., 2011). In Indonesian and Chinese contexts, female students also showed a more positive attitude towards science than male students (Aini et al., 2019; Liu et al., 2010). The other studies showed a different result that favours males (Hacieminoglu, 2016; Oon et al., 2020; Smith et al., 2014), claiming that male students tend to show better science performance, strengthening the stereotype that science is male-dominant. Meanwhile, the difference in societal expectations and the prominent learning environment gradually weaken science-gender stereotypes,

allowing male and female students to show a higher attitude equally. Several factors, such as cultural influence and social stereotypes, contribute further to our understanding of attitudes toward science.

The findings of this study confirm that there are no differences between male and female students in attitudes towards science, and both groups showed a moderate attitude. In the case of some attitude categories, female students showed a higher positive attitude than male students. However, these findings are explained with some limitations specific to secondary school in the Indonesian context. Moreover, the assessment of attitudes towards science with ATSQ focuses on the three-dimensional framework, explained by four categories. The additional latent variable in each dimension is suggested to obtain a more comprehensive interpretation of attitudes towards science. The study in the assessment of attitude is recommended with a different context of studies. In addition, further investigation of other factors that potentially influence the attitude towards science is suggested.

The present study used the ATSQ with good validity and reliable results to assess attitudes toward science in Indonesia. The result showed that the students have positive attitudes towards science indicated by the logit value above 0. There are no differences between male and female in their overall attitudes toward science, thus rejecting the hypothesis. However, in some categories (ENJ, PAR, and SVAL), female students have higher scores than male students; conversely, in the ANX category, male students have higher scores than female, indicating a more positive attitude among female students. Indeed, these results should be interpreted in light of limitations related to the context of the study. Therefore, more studies are suggested to conduct similar research in different contexts. In addition, a study of factors that can influence attitudes toward science is also recommended.

## **5.6 The assessment of science motivation**

This part of study aimed to measure students' science motivation in Indonesia. We used SMQ II for science motivation measurement with adaptation to the Indonesian language. Prior to the assessment process, a psychometric analysis was performed to verify the validity and reliability of the questionnaire, continuing with the analysis of students' profile in science motivation. The sixth hypothesis (H6) was tested in this subchapter, predicting that there is a significant difference between gender in science motivation, when females outperforming males (Oppermann et al., 2021; Schürmann & Quaiser-Pohl, 2022).

### 5.6.1 The validation of Science Motivation Questionnaire

The Science Motivation Questionnaire (SMQ II) showed high reliability (Cronbach's  $\alpha = .949$  and McDonald's  $\omega = .949$ ) that confirmed the internal consistency of the questionnaire. For the construct validity analysis with CFA. The result indicated that the 5-factor model is acceptable by the good fit indices ( $\chi^2/df = 10.23$ , CFI = .94, TLI = .94, RMSEA = .08, and SRMR = .04). The factor loading for each factor ranges between 0.63–0.88 meaning that the item distribution is fit to the position of each factor. Further analysis was done to ensure the quality of each item with the Rasch model analysis (Table 21). The analysis results in an acceptable fit value for both the unidimensional model (infit MNSQ = 0.73–1.61) and multidimension model (infit MNSQ = 0.82–1.66) model. Therefore, one item (SM02) has an out limit infit MNSQ value. However, we consider this item to remain in the questionnaire if the value does not exceed 1.7 (Bond & Fox, 2013). The point measure correlation (PTMA) also indicated a positive value that supports the fitting of each item in the construct.

Table 21. The item fit analysis results for the SMQ II

Variable	Item	Total score	Unidimension			Multidimension			PTMA
			Measure	SE	Infit MNSQ	Measure	SE	Infit MNSQ	
Intrinsic motivation (IM)	SM01	4196	0.46	0.04	1.09	0.18	0.04	1.23	0.58
	SM03	4192	0.47	0.04	0.89	0.19	0.04	0.99	0.67
	SM12	4112	0.62	0.04	0.93	0.39	0.04	1.10	0.67
	SM17	4662	-0.46	0.04	1.30	-0.93	0.04	1.45	0.57
	SM19	4205	0.44	0.04	0.87	0.16	0.07	0.94	0.68
Career motivation (CM)	SM07	4544	-0.23	0.04	1.02	-0.34	0.04	0.99	0.67
	SM10	4523	-0.19	0.04	0.90	-0.29	0.04	0.83	0.70
	SM13	4479	-0.10	0.04	0.98	-0.18	0.04	0.91	0.70
	SM23	4238	0.38	0.04	1.25	0.40	0.04	1.23	0.66
	SM25	4233	0.39	0.04	0.91	0.41	0.07	1.06	0.69
Self-determination (SD)	SM05	4711	-0.56	0.04	0.92	-0.89	0.04	1.01	0.69
	SM06	4157	0.53	0.04	1.12	0.43	0.04	1.17	0.62
	SM11	3816	1.19	0.04	1.00	1.24	0.03	1.13	0.64
	SM16	4440	-0.02	0.04	0.82	-0.24	0.04	0.91	0.69
	SM22	4561	-0.26	0.04	0.86	-0.53	0.07	0.92	0.69
Self-efficacy (SE)	SM09	4169	0.51	0.04	1.01	0.29	0.04	1.01	0.63
	SM14	4299	0.26	0.04	0.78	-0.02	0.04	0.82	0.69
	SM15	4248	0.36	0.04	0.73	0.10	0.04	0.78	0.72
	SM18	4307	0.24	0.04	1.16	-0.04	0.04	1.12	0.63
	SM21	4424	0.01	0.04	0.75	-0.32	0.07	0.85	0.70
	SM02	4873	-0.89	0.05	1.61	-0.10	0.04	1.66	0.54
	SM04	4933	-1.01	0.05	1.07	-0.25	0.04	0.99	0.67

Grade	SM08	4711	-0.56	0.04	0.99	0.31	0.04	1.01	0.71
motication	SM20	4796	-0.73	0.04	0.95	0.10	0.04	1.05	0.65
(GM)	SM24	4857	-0.85	0.05	1.00	-0.06	0.07	1.03	0.68

The item distribution level was determined by its logit measure, which is illustrated in the Wright map (Figure 38). In the uni- or multidimensional model, item SM11 showed the highest logit measure (1.19/1.24) indicating that students tend to give lower scores to respond to this item. In addition, item SM04 has lowest logit measure (-1.01) compared to the other items, meaning that students give higher scores to respond to these items.

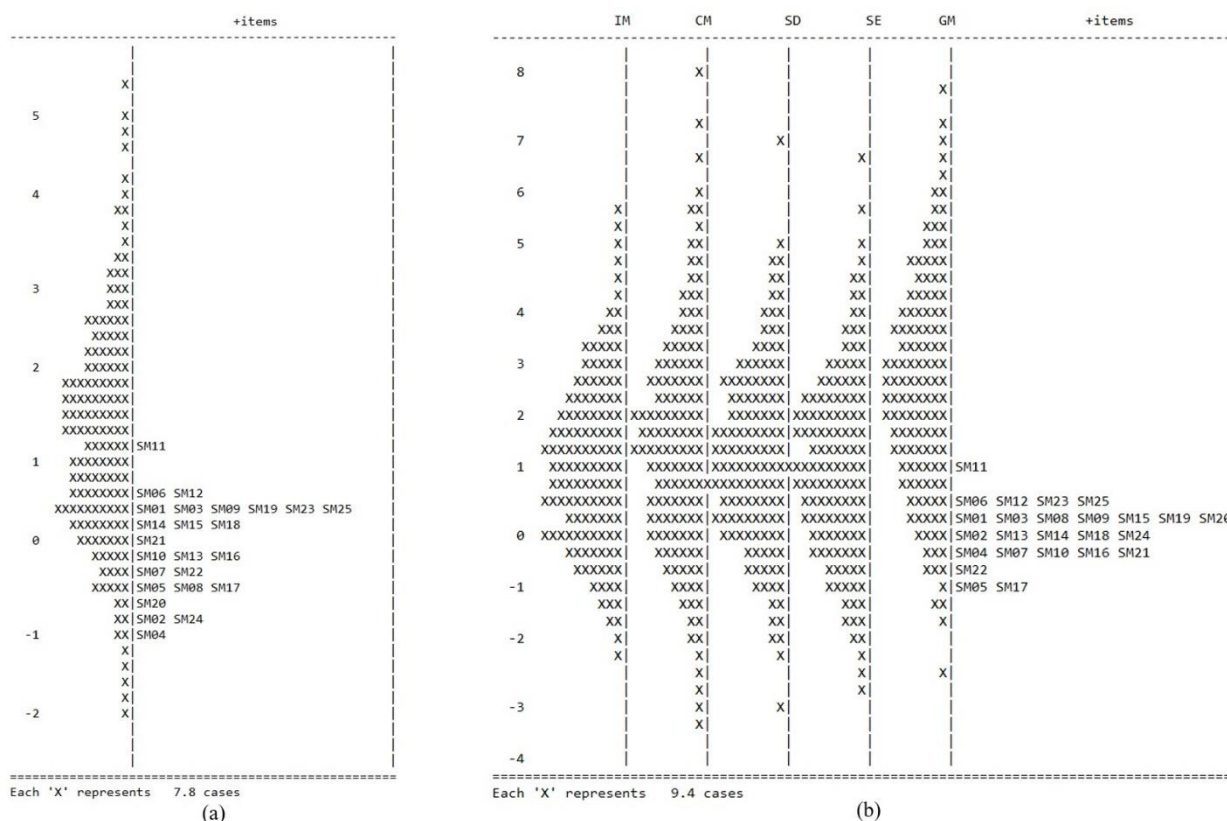


Figure 38. The Wright map of the SMQ II items (a = unidimension, X represents 7.8 cases; and b = multidimension, X represents 9.4 cases)

The analysis of the item bias was done concerning the gender group using different item functions (DIF). The DIF result showed that seven out of 25 items in the science motivation questionnaire have a significant result (items SM03, SM08, SM09, SM17, SM20, SM23, and SM24). Thus, further identification with DIF size examination was done to measure the size effect of DIF (Figure 39). Among those seven significant items, the size was small ( $< 0.64$ ) which can be explained that the bias effect is low or negligible.

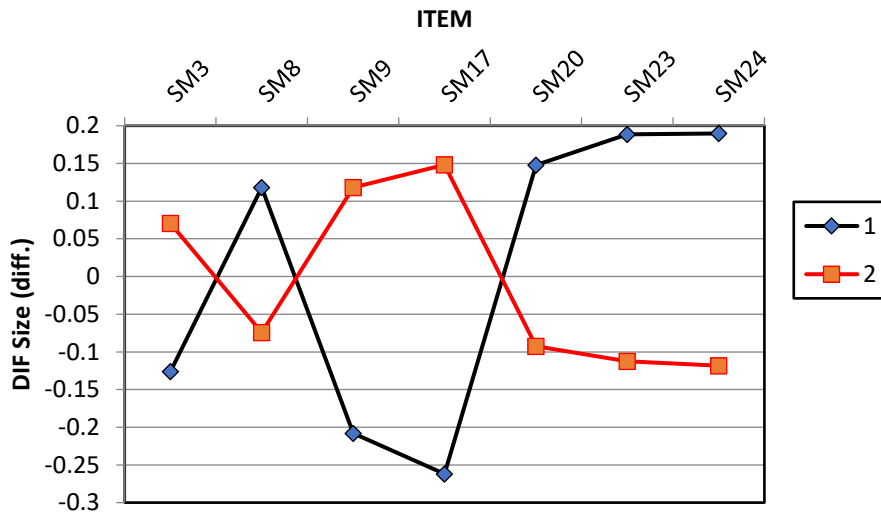


Figure 39. The DIF results for SMQ II (1 = male; 2 = female)

### 5.6.2 The profile of students' science motivation

The distribution of the students' science motivation is presented in Figure 40. The mean student score for SMQ II is 1.26 logit (SD = 1.50). Among the five variables in SMQ II, GM has the highest average logit (M = 2.43; SD = 2.30), followed by CM (M = 1.76; SD = 2.72), SE (M = 1.70; SD = 2.80), SD (M = 1.64; SD = 2.36), and the last is IM (M = 1.21; SD = 2.03). The students' overall responses in SMQ II show a value above average 0 logit, indicating that their science motivation level is approximately high. In addition, the distribution of students' score in SMQ II has a skewness value of 0.49 and a kurtosis value of 2.38, showing the criteria of normal distribution. In each SMQ II category, students' responses also followed a normal distribution (skewness = 0.13–0.24; kurtosis = -0.47–0.56).

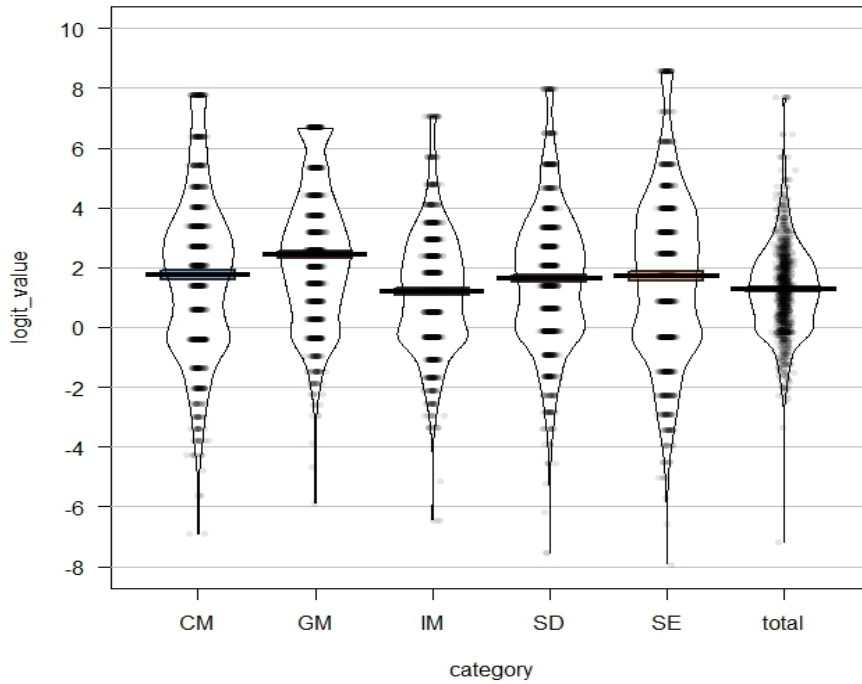


Figure 40. The distribution of students' logit scores in science motivation (CM = career motivation; GM = grade motivation; IM = intrinsic motivation; SD = self-determination; SE = self-efficacy)

### 5.6.3 The gender difference in science motivation

The distribution of the science motivation of male and female students is shown in Figure 41. When comparing science motivation between male and female students, we found a significant difference between both genders ( $M_{\text{male}} = -0.19$ ,  $SD_{\text{male}} = 0.76$ ;  $M_{\text{female}} = 0.21$ ,  $SD_{\text{female}} = 0.58$ ;  $t = -0.66$ ,  $p < .05$ ). Among SMQ II variables, there was a significant difference is found in the CM, IM, SD, SE and GM variables when female students showed a better score than male students (see Table 22), indicating that female students have more positive motivation rather than male students.

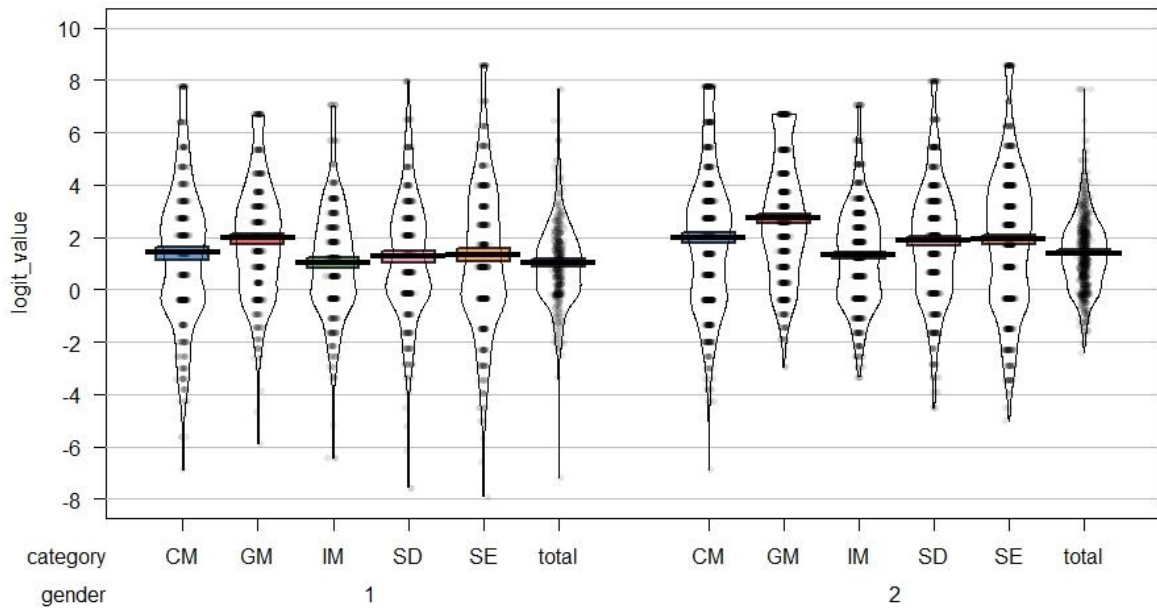


Figure 41. The distribution of students' science motivation (CM = career motivation; GM = grade motivation; IM = intrinsic motivation; SD = self-determination; SE = self-efficacy) by gender (1 = male; 2 = female)

Table 22. The gender comparison of students' science motivation

Category	Male (n = 457)			Female (n = 786)			t-test (df = 1241)	Cohens' d
	M <sub>score</sub>	M <sub>logit</sub>	SD	M <sub>score</sub>	M <sub>logit</sub>	SD		
IM	3.38	1.04	2.13	3.47	1.31	1.96	t = -2.24, p = .026*	-0.14
CM	3.44	1.40	2.76	3.60	1.98	2.68	t = -3.60, p < .001**	-0.22
SD	3.38	1.26	2.38	3.55	1.86	2.31	t = -4.32, p < .001**	-0.26
SE	3.36	1.32	2.89	3.50	1.91	2.73	t = -3.61, p < .001**	-0.21
GM	3.74	1.96	2.25	3.97	2.71	2.30	t = -5.60, p < .001**	-0.32
Total	3.46	1.01	1.58	3.62	1.39	1.43	t = -4.31, p < .001**	-0.28

Note: CM = career motivation; GM = grade motivation; IM = intrinsic motivation; SD = self-determination; SE = self-efficacy. \*\* significant at  $p < .001$ , \*  $p < .05$

#### 5.6.4 Discussion

The Science Motivation Questionnaire (SMQ II) used in this study showed an acceptable validity and reliability result with the Indonesian sample. The SMQ II with five factors is explained well within 25 items and shows no bias in assessing students' science motivation. Consistent with this result, SMQ II has also been shown to be valid for motivation measurement with different context studies (Glynn et al., 2011; Schumm & Bogner, 2016; You et al., 2018). Thus, this questionnaire is recommended for further studies in science motivation.

The results in a high level of science motivation, indicating that students in Indonesia express positive motivation related to science subject and learning. Among the five variables in science motivation, students have a higher score in grade and career motivation, followed by self-efficacy, self-determination, and intrinsic motivation. From this point on, students' motivation is predominantly driven by external factors regarding grades and future science-related careers. This result is in line with previous studies that students are motivated to learn science mainly because of the grades (Campos-Sánchez et al., 2014; Glynn et al., 2011; Schumm & Bogner, 2016). Vedder-Weiss & Fortus (2012) also claimed that external learning goals have a greater impact on student motivation than their internal goals. Similarly, students who are interested in a science-related career will put more effort into participating in science learning and show more motivation in learning activities. Kenny et al. (2006) explained that students with clear career plans and positive expectations for their career success may express more feelings in valuing course learning and school activities. The implication of these findings can be explained by the involvement of grades and rewards, as well as a precise career projection in the science learning process. In addition, internal motivation also plays an important role in student learning success. It needed to be nurtured along with external feedback through a meaningful learning process. Indeed, the learning process should emphasise curiosity and challenge to make students participate and engage in all learning activities (Carreira, 2011).

As regards gender comparison, there is a significant difference between male and female students in science motivation, thus supporting the hypothesis (H6). Female students showed more motivation for science than male students. However, male students showed more diversity in their responses, implying that some male students have the highest motivation and other male students have the lowest motivation in science. The favours of female towards male students in science motivation are also supported by several studies (Oppermann et al., 2021; Schürmann & Quaiser-Pohl, 2022). Cabras et al. (2023) confirm that female students showed more positive intrinsic and extrinsic motivation than male students in the Italian and Russian samples. Other studies also reported that male students prefer to do science activities outside of school and are less motivated during science learning (Ozbas, 2016; Salmi & Thuneberg, 2019). Female students also indicated that they have more awareness of themselves as learners and consider that science is important for their future careers, which drives their motivation to learn science (Daher et al., 2021). Therefore, cultural aspects and gender stereotypes in science can be considered possible explanations for gender differences in science motivation (Cabras et al., 2023).

The present study confirms that students have a high level of science motivation with favour of females towards males in each category of science motivation. This result has some limitations, as the measurements of science motivation with SMQ II addressed mostly intrinsic and extrinsic motivation, as well as self-efficacy and self-determination. Other theoretical frameworks and measurements for science motivation are also available to measure science motivation (e.g. Dermitzaki et al., 2013; Tuan et al., 2005). Further analysis related to gender stereotypes and cultural aspects of the sample context may be necessary to provide a more detailed explanation of the differences between male and female students in science motivation. Therefore, the result of this scientific motivation assessment may be useful for further educational assessment and training programmes in Indonesia and as a comparison with different studies in the other context.

The SMQ II is confirmed as a valid and reliable tool to measure students' science motivation in Indonesia. The results of the assessment demonstrate positive results of students' science motivation. The highest score was found in the grade and the motivation for the career represented extrinsic motivation. Therefore, the motivation of the students is mostly driven by extrinsic rather than intrinsic aspects. Emphasizing external factors in learning practices is recommended to improve students' motivation, whereas focusing on internal factors through meaningful learning is also necessary to raise student motivation even further. The present study supports the hypothesis that shows gender differences in science motivation, favouring female students over male students. The explanation of this result could be connected to cultural aspects and gender stereotypes in science learning. Furthermore, the identification of several factors connected to science motivation is recommended for future studies.

### **5.7 The relationship between cognitive and noncognitive factors on problem solving**

In this part of the study, we analyse the relationship between domain-general and domain-specific problem-solving and its connection with cognitive factors (inductive reasoning and scientific competency), and non-cognitive factors that include affective factors (attitude towards science and science motivation), gender, and socio-economic status. As the theoretical foundation for connecting these factors to problem solving has been established (see Chapter 2), we offer empirical evidence confirming the effects of these factors on problem-solving abilities. Thus, we hypothesize that domain-general and domain-specific problem solving is predicted by inductive reasoning (H7), scientific competency (H8), attitude (H9), motivation (H10), gender (H11), and socioeconomic status (H12).

### 5.7.1 The result of empirical analysis

The correlation analysis was performed and indicated that MicroDYN is significantly correlated with gender ( $r = -.17, p < .01$ ), SES ( $r = .16, p < .01$ ), SC ( $r = .40, p < .01$ ), IR ( $r = 0.38, p < .01$ ), ATSQ ( $r = .09, p < .01$ ), and SMQ II ( $r = .07, p < .01$ ). While SPS also significantly correlates with SES ( $r = .08, p < .01$ ), SC ( $r = .38, p < .01$ ), IR ( $r = .32, p < .01$ ), ATSQ ( $r = .08, p < .01$ ), and SMQ II ( $r = .10, p < .01$ ) (see Table 23). Through the correlation analysis, we see that cognitive factors have a medium correlation with domain-general and domain-specific problem-solving, while the relationship with affective factors is weak. In addition, both cognitive and affective factors have a significant correlation with each other. Other variables, such as SES and gender have a weak connection to problem solving, cognitive and affective factors.

Table 23. The correlation between gender, socioeconomic status, cognitive and affective factors, and problem solving

Variable	1	2	3	4	5	6	7	8
1. Gender	1							
2. SES	-.06*	1						
3. MicroDYN	-.17**	.16**	1					
4. SPS	.05	.08**	.31**	1				
5. SC	.17	.16**	.40**	.38**	1			
6. IR	-.19	.21**	.38**	.32**	.59**	1		
7. ATSQ	.04	.05	.09**	.08**	.10**	.12**	1	
8. SMQ II	.13**	.05	.08**	.10**	.12**	.15**	.75**	1
$M_{score}$	-	3.39	0.11	0.35	0.48	0.62	3.15	3.56
(SD)	-	(1.24)	(0.15)	(0.24)	(0.24)	(0.22)	(0.39)	(0.57)
$M_{logit}$	-	0.90	-3.93	-0.88	-0.16	0.69	0.21	1.26
(SD)	-	(2.79)	(1.80)	(1.43)	(1.34)	(1.41)	(0.65)	(1.50)

Note: SES = socioeconomic status; SPS = science problem solving; SC = scientific competency; IR = inductive reasoning; ATSQ = Attitude towards Science Questionnaire; SMQ = Science Motivation Questionnaire.

\*significant at  $p < .05$ ; \*\*  $p < .01$

Furthermore, a SEM analysis was performed to confirm the effect of each variable on domain-general and domain-specific problem-solving. The proposed model produces an acceptable fit ( $\chi^2/df = 6.42$ , CFI = .92, TLI = .90, RMSEA = .07, and SRMR = .04). As indicated in Figure 42 and Table 24, science problem-solving is associated with MicroDYN ( $\beta = .21; p < .01$ ). In the case of cognitive factors, we found that inductive reasoning directly affects MicroDYN ( $\beta = .20; p < .01$ ) and strongly affects scientific competency ( $\beta = .75; p < .01$ ), there is no significant direct effect of inductive reasoning on science problem-solving. Furthermore, scientific competency is the only cognitive factors that has direct affects to science

problem-solving ( $\beta = .34; p < .01$ ). Scientific competence was also significantly associated with MicroDYN ( $\beta = .35; p < .01$ ).

In affective factors, attitudes towards science and science motivation have a direct effect on MicroDYN ( $\beta = .29; p < .05; \beta = -.24; p < .05$ ). Regarding gender, we found its significant effects on MicroDYN ( $\beta = -.19; p < .01$ ) and science problem-solving ( $\beta = .09; p < .01$ ). There is no effect of gender on the cognitive factor, but it has a small effect on attitude towards science ( $\beta = .09; p < .01$ ) and science motivation ( $\beta = .13; p < .01$ ). We found no direct effect of socio-economic status on problem solving. This variable is associated only with affective factors and inductive reasoning ( $\beta = .26; p < .01$ ).

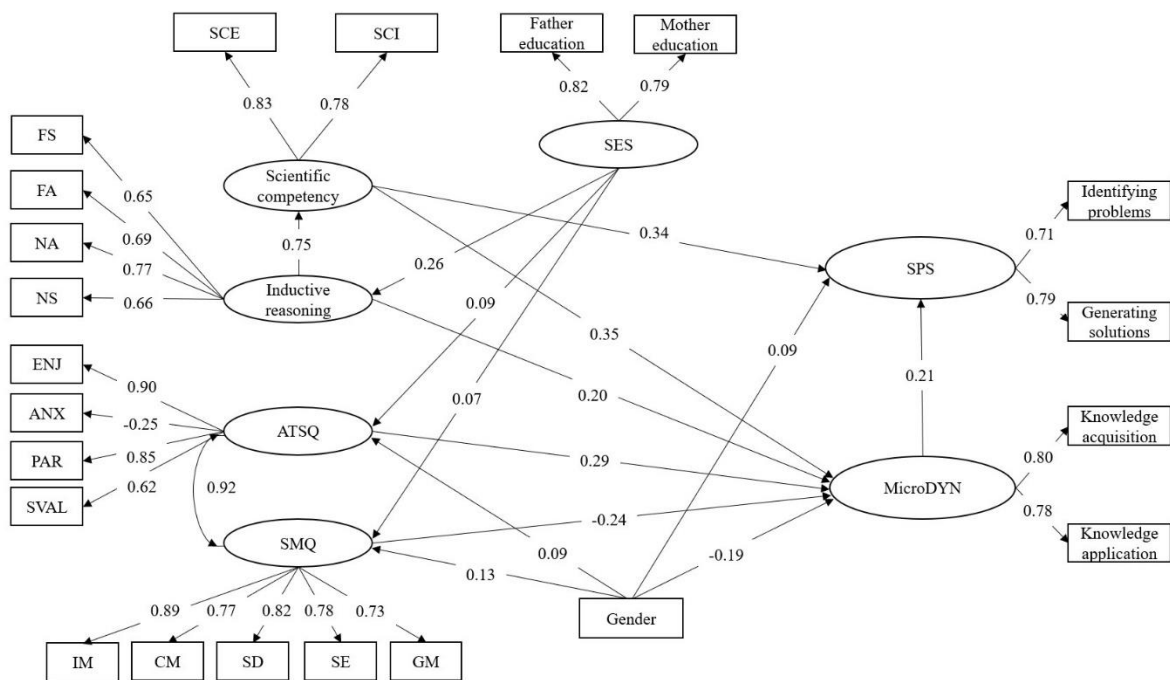


Figure 42. The relationship model between cognitive and noncognitive factors on problem solving (Note. SES = socioeconomic status; SPS = science problem solving; ATSQ = Attitude towards Science Questionnaire; SMQ = Science Motivation Questionnaire)

Table 24. Standardised results of model analysis

Model	$\beta$	SE	<i>P</i>
gender → science problem-solving	.09	.03	.007
gender → MicroDYN	-.19	.03	.000
gender → science motivation	.13	.02	.000
gender → attitude towards science	.09	.03	.003
socio-economic status → inductive reasoning	.26	.03	.000
socio-economic status → attitude towards science	.09	.03	.008
socio-economic status → science motivation	.07	.03	.025
MicroDYN → science problem-solving	.21	.05	.000
scientific competency → science problem-solving	.34	.07	.000
scientific competency → MicroDYN	.34	.06	.000
inductive reasoning → MicroDYN	.20	.06	.001
inductive reasoning → scientific competency	.75	.02	.000
attitude towards science → MicroDYN	.29	.11	.011
science motivation → MicroDYN	-.24	.11	.032

Note:  $\beta$  = standardised coefficient; SE = standard error

### 5.7.2 Discussion

The present study concludes the positive connection between domain-general and domain-specific problem-solving. Since both problem solving require similar strategies to deal with each problem situation (Molnár et al., 2013), domain-general problem-solving provides knowledge acquisition and application strategies that are useful for students to solve domain-specific problems. The only difference is that domain-general problem-solving has less depending on prior knowledge, whereas domain-specific problem-solving benefits from the knowledge acquisition strategy and requires support from prior knowledge. Indeed, domain-specific problem-solving is mostly based on knowledge application that makes students should utilise their knowledge even before solving the problem. Therefore, domain-general problem-solving requires students to gain new knowledge and apply it in a new situation. As the regression coefficient from domain-general to domain-specific problem solving is significant, fostering domain-general problem solving through an educational programme will enhance students' domain-specific problem-solving ability.

#### *The role of cognitive factors on problem solving*

In cognitive factors, inductive reasoning has a significant effect on domain-general problem-solving, but not on domain-specific problem-solving, thus rejecting hypothesis (H7). Meanwhile, scientific competency has a significant effect on both domain-general and domain-specific problem-solving which support the hypothesis (H8). The effect of inductive reasoning

on domain-general problem-solving has been confirmed by many studies (Molnár et al., 2013; Stadler et al., 2015). Inductive reasoning mostly influences domain-general problem-solving during knowledge acquisition and problem identification. In this phase, students explore the system by manipulating and controlling variables to identify the main effect, multiple effects, and multiple dependence of variables and find their connection, which serves as new knowledge to help them understand the problems. The knowledge acquisition phase also requires students to observe situations, draw inferences, and identify causes to reach unexpected findings, which employ several cognitive strategies such as causal, deductive, and inductive reasoning (Dunbar, 2000). Additionally, the nonsignificant effect of inductive reasoning to domain-specific problem-solving could be explained that the role of inductive process is not directly connected to well-defined science problem solving, where the information of problem situation strongly rely on domain knowledge. Indeed, the role of inductive reasoning potentially enhance the acquisition and reconstruction of individual knowledge, meaning that the influence of inductive reasoning to domain-specific problem-solving is supported by domain knowledge. This suggestion raises due to the finding that inductive reasoning strongly connected to scientific competency, where scientific competency positively predicts domain-specific problem-solving. Thus, we assumed that the influence of inductive reasoning to domain-specific problem-solving is mediated by scientific competency.

On the other hand, the influence of scientific competency on domain-general and domain-specific problem solving can be explained by the structure of scientific competency, which includes knowledge and cognitive processes. Several types of knowledge (i.e. factual, conceptual, and procedural) are required to solve problems (Mayer & Wittrock, 2006). When students encounter a specific problem that requires a high interpretation of problem situations, their prior knowledge in that context is beneficial for analysing and generating new ideas. Scientific competency encompasses the ability to explain phenomena and to interpret and evaluate data or evidence, which is applied during both domain-general and domain-specific problem-solving tasks. As seen during the knowledge acquisition phase, students explore the system by manipulating input variables and identifying changes in output variables. They deal with some information and data that reveal the causal relationship between input and output variables. Scientific competency will foster this stage by providing cognitive processes to evaluate and interpret data during the system exploration. Indeed, successful system exploration will lead students to find the rule and draw a conclusion. In domain-specific problem-solving, scientific competency will be useful during problem identification and solution generation. The basic knowledge that students acquire during learning activities is reported to support them in

applying problem-solving strategies (Hambrick & Engle, 2003; She et al., 2012). The other study also found that problem-solving ability will also increase along with their domain-specific knowledge achievement (She et al., 2012). Vieira & Tenreiro-Vieira (2016) also confirmed that students' ability in science supports them in making judgements when faced with a challenge, problem, or issue by justifying claims with accurate evidence and logical criteria.

#### *The role of affective factors on problem solving*

As for affective factors, we found that attitude towards science and science motivation have significant effects on domain-general problem solving, but not on domain-specific problem solving, thus rejecting the hypothesis (H9 and H10). This result is inconsistent with other studies, which showed a positive correlation between attitude and motivation and domain-specific problem solving (Baars et al., 2017; Guven & Cabakcor, 2013; Marchis, 2013). Similarly, Zakaria & Ngah (2011) also found no significant correlation between attitude and domain-specific problem-solving abilities, explaining that high perseverance and willingness in solving problems do not strongly support the initial skills and abilities in problem-solving. However, we found a positive connection between attitude and motivation to domain-general problem-solving. This finding can be explained by the role of attitude in connection to the learning activities and the use of strategies to effectively solve a problem. Attitude correlates with learning strategies that can potentially improve problem-solving achievement (Luo et al., 2019). Taghizadeh & Hajhosseini (2021) found that attitude contributes to driving students to be more active and satisfied with their learning. The more students actively engage in learning, the greater the chance of improving their problem-solving abilities, depending on how problem-solving strategies are applied during learning. Additionally, emotions as part of attitude constructs can alter and initiate individual cognitive processing. It influences how information can be stored and retrieved (Kim & Pekrun, 2014). People with positive attitudes use their general knowledge more effectively to solve problems than those with negative attitudes. Negative attitude has an impact on reducing information processing (Pekrun, 2006). Similar to attitude, motivation also supports problem solving by its connection to student learning performance (Baars et al., 2017; Bereby-Meyer & Kaplan, 2005). Motivation sustains student behaviour and allows them to engage and be active in many tasks (Fredricks et al., 2004). Motivation is also related to persistence when students face obstacles during problem-solving activities, they tend to overcome them. From this perspective, students with higher motivation are eager to engage in every stage of problem-solving tasks and demonstrate greater persistence in achieving their goals.

### *The role of gender and socioeconomic status on problem solving*

Regarding gender, this study found the effect of gender on domain-general and domain-specific problem-solving, thus supporting the hypothesis (H11). Recent studies of gender differences in problem solving have been conducted with favour males toward females (OECD, 2014; Wüstenberg et al., 2014), supporting the theoretical studies in cognitive studies that males have an advantage in spatial and abstract information processing over females (Halpern & LaMay, 2000; Reilly & Neumann, 2013). However, Borgonovi & Greiff (2020) found that the connection between gender and problem solving varied in several countries. In countries with high gender inequality, males tend to outperform females, while some countries showed no or small gender gaps in problem-solving performance. These findings conclude that gender differences are also related to gender equality and sociocultural aspects. Hence, even if males possess a greater cognitive potential that benefits them in problem solving, the influence of cultural aspects related to the learning process and environment can be a potential factor in explaining gender differences in problem-solving ability. In addition, we found no direct effect of socioeconomic status on problem solving, thus rejecting hypothesis (12). This finding is supported by several studies that have confirmed the nonsignificant effect of socio-economic status concerning the education level of both father and mother (Amalina & Vidákovich, 2023b; Guven & Cabakcor, 2013). Instead, socioeconomic status is connected to inductive reasoning, attitude towards science, and science motivation, which could lead to the possibility of mediation effect of these variables in the relationship between socioeconomic status and problem solving.

### *Summary*

This study showed that domain-general problem-solving is directly affected by gender and cognitive and affective factors, while domain-specific problem solving is directly affected by gender and scientific competency. Furthermore, socio-economic status does not have a direct effect on problem solving but is significantly related to inductive reasoning and affective factors. This study suggests that improving students' problem-solving ability also requires fostering cognitive and affective factors in educational practices. Therefore, the problem-solving programme should consider the integration of cognitive aspects and maintain the affective needs of students during classroom intervention and other nonformal activities. Furthermore, the effect of cognitive and affective factors on problem solving is limited to direct effect calculation. Further analysis of the mediation effect between variables and the involvement of other educational and sociocultural variables may be necessary to gain a more comprehensive

understanding of problem-solving dynamics in education practices. Therefore, further research is recommended to be done in the near future with the inclusion of other predictor and mediator variables to explain problem-solving ability.

In the current study, we found a direct effect of gender, cognitive factors (inductive reasoning and scientific competency), and affective factors (attitude towards science and science motivation) on domain-general problem-solving. Meanwhile, domain-specific problem solving is directly affected by domain-general problem solving, scientific competency, and gender. Socio-economic status has a direct effect on inductive reasoning and affective factors, but no direct effect on problem solving, which leads to a possibility of a mediation effect in its connection to problem solving. Hence, further research on mediation effect of several variables on problem solving is suggested, along with the exploration of other educational and socio-cultural factors that potentially influence problem-solving.

## **CHAPTER 6 GENERAL CONCLUSION, IMPLICATION, LIMITATION, AND RECOMMENDATION**

The dissertation provides information about the assessment of problem solving in the Indonesian context. Based on the literature review, there is a need for improvement in the assessment of problem solving in Indonesia, as recent studies have focused mainly on problem-solving interventions, but there is a lack of research in the area of assessment. The current research on problem-solving assessment in Indonesia has been reported to be insufficient to generate information on student performance in problem solving. The existing measurement is only domain-specific and has less comprehensive with psychometrical properties (Wicaksono & Korom, 2022b). To this extent, in this study, the assessment of problem solving in general and specific domain was carried out with the validity evidence of its measurement. In problem-solving tasks, some sequences and activities that require cognitive processes and mental representation of the test taker (Greiff et al., 2012). Indeed, the involvement of cognitive factors as well as affective factors, gender, and socioeconomic background that are connected to individual differences is also important to get better understanding of phenomena of problem solving. This study also investigates the influence of cognitive and affective factors, gender and socioeconomic status on problem solving.

### **6.1 General conclusion**

The present study explains the assessment results of domain-general and domain-specific problem solving. In addition, the measurement of cognitive factors such as inductive reasoning (IR) and scientific competency (SC), affective factors in attitude towards science (ATSQ) and science motivation (SMQ II) was done. The prior analysis of the validity of all measurements confirms that the tests and questionnaire used in this study are valid and reliable for evaluation purposes. Each test and questionnaire show acceptable factor analysis with proficient result in each dimension variable. Rasch analysis also confirms that the items in the tests and questionnaire are in the range of acceptable results (infit MNSQ 0.5–1.5). We find no or negligible risk of item bias in the instrument through DIF analysis concerning gender. The psychometric properties of the instruments show valid evidence and are sufficient to be applied for assessment purposes in the Indonesian context.

The assessment results of domain-general and domain-specific problem-solving follow descriptive analysis based on logit parameters. Student performance in both domain-general and domain-specific problem solving was classified as low level with an average logit value

below 0 point. In domain-general problem solving, two phases are used for the assessment, knowledge acquisition and knowledge application. Students also show low performance in knowledge acquisition and knowledge application, indicating the difficulty of students in generating knowledge from problem exploration activities with low chance to apply their knowledge in the new situation. The gender differences are found with males significantly outperforming female students in domain-general problem solving as well as in knowledge acquisition and knowledge application phase, supporting the hypothesis (H1). This finding corresponds to other studies that favour males towards females in domain-general problem solving (Greiff et al., 2018; OECD, 2014; Wüstenberg et al., 2014). In domain-specific problem solving, the assessment also employs two phases, identifying problem (IP) and generating solution (GS) that present problem situations in science phenomena. The students show low performance in the IP and GS categories of domain-specific problem solving. In contrast to domain-general problem solving, the comparison result does not indicate a significant difference between domain-specific problem solving of male and female students, thus rejecting the second hypothesis (H2). Although some studies favoured males towards females in problem solving (Gok, 2014; Soto-Ardila et al., 2022), many studies also found no gender differences in maths problem solving (Gallagher et al., 2000; Guven & Cabakcor, 2013) and science problem solving (Harskamp et al., 2008). Since prior knowledge is required during the completion of domain-specific problem-solving tasks, the similar acquisition of student knowledge during learning and intervention at school can potentially influence their performance in domain-specific problem solving.

In cognitive factor, the measurement of inductive reasoning reveals that students achieve a score above average, higher than 0 logit. Among the four subtests of the inductive reasoning test, the students got the highest score in number series and the lowest score in number analogy. The students' performance is more varied when they solve series tasks rather than analogy tasks, which is confirmed by higher standard deviation. Students have a similar ability to make analogies, but when it comes to making order and finding the pattern of attributes, their ability is diverse. This study also shows no gender differences in inductive reasoning tasks, resulting in rejection of the hypothesis (H3). This result is supported by previous studies that found similar performance between male and female students in inductive reasoning (Díaz-Morales & Escribano, 2013; Kambeyo, 2018; Molnár, 2011; Van Vo & Csapó, 2020). The difference is only found in the number analogy that favours males towards females, but the effect size is relatively small. The other cognitive factor, the measurement of scientific competency, reveal that the students' scores are close to 0 logit on average, presenting a moderate level of scientific

competency. In each category of scientific competency, students show better results in explaining scientific phenomena (SCE) than in interpreting and evaluating scientific data or evidence (SCI). Indeed, SCI requires a higher cognitive ability to evaluate and analyse compared to SCE which employs the ability to understand and recall information. In a gender comparison between the ability of male and female students in scientific competency, we found no significant differences and rejected the hypothesis (H4). Gender differences in connection with scientific competency are varied as a result of the context and the level of gender equity influences (Cheng et al., 2021). Indeed, the science competency test is built based on the science education curriculum; this nonsignificant result can be explained by similar cultural context and learning experience throughout their education.

In affective factors, the students' attitude towards science is presented on a moderate level with the highest score in the value of science. This result indicates a higher understanding and acceptance of the influence of science in personal life and communities. The lowest attitude category is found in the anxiety and difficulty variable, indicating that the students perceived fewer negative emotions during science learning. The gender differences analysis does not produce a significant result in the overall attitude towards science of students, rejecting the hypothesis (H5). However, in each category, female students showed a better perception of enjoyment, value of science, and participation in science learning and activities, while male students showed higher anxiety and difficulty levels than female students. In the science motivation assessment, students get a high overall score (above 0 logit). Among five categories of science motivation, students show the highest score in grade motivation followed and career motivation, whereas the lowest score is in intrinsic motivation. This result implies that the motivation of the students is mainly driven by external factors such as grade and future career. In fact, students are motivated to learn science because they want to get high grades rather than their inherent satisfaction and willingness to learn science. Since the students also showed high intrinsic motivation, self-determination, and self-efficacy, this internal source of motivation also contributes to the learning behaviour of the students. Regarding the gender comparison, there is a significant difference between the science motivation of male and female students, supporting the hypothesis (H6). Female students show higher scores in each category of science motivation than male students, which concludes that female students are more motivated to learn science.

Furthermore, the results of this study confirm the effect of cognitive and affective factors, gender, and socio-economic status on problem solving. In cognitive factor, inductive reasoning has a significant direct effect only on domain-general problem solving, but not on domain-

specific problem solving (rejecting H7). Additionally, scientific competency has a direct effect on both domain-general and domain-specific problem solving (supporting H8). In the case of affective factors, attitude towards science and science motivation both have a direct effect on domain-general problem solving, but no direct effect on domain-specific problem solving (rejecting H9 and H10). In addition, gender has a direct effect on domain-general and domain-specific problem solving (supporting H11), as well as affective factors, while socio-economic status has no effect on problem solving (rejecting H12). The role of a cognitive factor in problem solving is explained as providing a cognitive basis to understand the problems, apply their knowledge, and make decisions to deal with the problem situation. Furthermore, affective factors will support students' activities and engagement, which potentially improve their strategies in dealing with problem-solving tasks. In fact, no significant effect of inductive reasoning and affective factor on domain-specific problem solving raises a concern about their indirect effect and the role of domain-general problem solving and scientific competency as a mediator to explain domain-specific problem solving.

## **6.2 Educational implication**

This study contributes to providing information on the level of problem-solving ability of students in Indonesia, as well as their level of cognitive and affective state. Additionally, valid measurements for problem-solving and other assessments can be implemented in educational practices in Indonesia. Specifically, while most schools use domain-specific problem-solving tests for learning evaluation purposes, introducing domain-general problem-solving assessments will benefit students. This approach allows students to engage with problem-solving tasks on topics that are familiar and related to their daily lives. As mentioned in the Indonesian national curriculum, problem solving is one of the educational goals for the high school level (BSNP, 2013). However, the result of this study suggests that Indonesian students in Grade 10 have low problem-solving performance. Several studies have recently been developed on learning and instruction in problem solving in Indonesia. Learning activities based on problem orientation and discovery have been suggested in the curriculum. Thus, the teachers and stakeholders must be aware of the details in the implementation of problem-based instruction with a focus on the problem-solving strategies that students can learn during classroom activities. The details in the learning process such as activities to acquire new knowledge and apply it in new situations should be emphasized during problem-solving strategies. It is also important to integrate inductive process into learning activities with connection to content knowledge from each subject will contribute to students' effectiveness in

applying problem-solving strategies (see Alibali et al., 2009; Herrmann et al., 2023; Tomic, 1995). In addition, maintaining students' attitude and motivation during learning activities is needed to foster problem-solving activities. Meaningful learning and a pleasant environment should be created during learning activities. It is also important to have more work on the assessment process that can give information about student progress and support further educational improvement and methodological efforts in Indonesia focused on problem solving.

Furthermore, this study contributes to the field of global educational research through its methodological and theoretical implications. The instruments presented in this study for assessing problem solving (MicroDYN and SPS), cognitive factors (IR and SC), and affective factors (ATSQ and SMQ) showed strong validity and reliability results. The domain-specific problem-solving framework can be adapted to different disciplines beyond the natural sciences. The domain-general problem solving, inductive reasoning, attitude, and motivation measurement addresses a broader context that can be applied to the assessment process in different samples and grade levels. These results also support the theoretical foundation regarding the role of cognitive and affective factors in problem solving empirically, which can serve as references for further research of problem solving. Additionally, the study on gender and socio-economic status will provide more evidence on problem solving from the perspective of individual differences.

### **6.3 Limitation and future recommendation**

The result of the study should be explained in light of limitations. The measurement of domain-specific problem-solving and scientific competency was done within the scope of a natural sciences topic, which may potentially show different results when applied to other disciplines. The framework for domain-specific problem solving was developed with a focus on two main phases of problem solving, while the other frameworks encompassed various phases (Rausch & Wuttke, 2016; Walker et al., 2016). Similarly, in the case of the scientific competency test, the framework focused on two main categories that do not fully represent all concepts of scientific competency. The other issue is connected to the inductive reasoning test, which is limited to series and analogy in both figural and numerical form. Other types of inductive reasoning tasks are available for assessment purposes, such as matrix, verbal, and scheme (see de Koning et al., 2003; Ifenthaler & Seel, 2011). Using these types of tests can provide further information about inductive reasoning. During the adaptation of the questionnaire for attitude towards science and science motivation, the translation process was adjusted to language and cultural aspects in Indonesia, which led to some changes in the

questionnaires but still maintained their meaning and reliability. The other concern also arises from the sample distribution, which focused only on several areas in Indonesia. Although the sampling technique is maintained to represent the population, the sample is limited to Java, and students from other areas were not included in the study. Future research should consider the participation of students from other areas of Indonesia. To provide comprehensive information on education in Indonesia, a large-scale survey based on a broader sample is needed.

Several recommendations can be made in the light of the findings of the study. First, in domain-general and domain-specific problem-solving tests, adding easier tasks is necessary because the majority of items are at a difficult level. Further adaptation of the domain-specific problem-solving tasks is recommended in other disciplines based on the subjects and context of the studies. The involvement of various domains (e.g. science, mathematics, technology, and social studies) could provide a comprehensive profile of students' domain-specific problem-solving abilities, which is important for the improvement of educational programmes, curricula and teaching methods. Second, instruction and training based on the inductive reasoning process and the components of scientific competency are suggested as an alternative programme to foster students' problem-solving ability. This study involved several cognitive and noncognitive factors to explain problem-solving ability. Several variables have the potential to affect problem-solving ability, such as metacognition (Schoenfeld, 2016; Zhao et al., 2019), divergent thinking (He & Wong, 2021; Vincent et al., 2002), students' beliefs about the problem-solving process (Güven & Cabakcor, 2013; Öztürk & Güven, 2016), and others. Therefore, it is also recommended to conduct more research that incorporates these variables to explain the complexity and dynamic of problem solving.

## References

- Abeyssekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research & Development, 34*(1), 1–14. <https://doi.org/10.1080/07294360.2014.934336>
- Adigwe, J. C. (1992). Gender Differences in Chemical Problem Solving amongst Nigerian Students. *Research in Science & Technological Education, 10*(2), 187–201. <https://doi.org/10.1080/0263514920100206>
- Aini, R. Q., Rachmatullah, A., & Ha, M. (2019). Indonesian primary school and middle school students' attitudes toward science: focus on gender and academic level. *Journal of Baltic Science Education, 18*(5), 654–667. <https://doi.org/10.33225/jbse/19.18.654>
- Alibali, M. W., Phillips, K. M. O., & Fischer, A. D. (2009). Learning new problem-solving strategies leads to changes in problem representation. *Cognitive Development, 24*(2), 89–101. <https://doi.org/10.1016/j.cogdev.2008.12.005>
- Almasri, F., Hewapathirana, G. I., Ghaddar, F., Lee, N., & Ibrahim, B. (2021). Measuring attitudes towards biology major and non-major: Effect of students' gender, group composition, and learning environment. *PLOS ONE, 16*(5), e0251453. <https://doi.org/10.1371/journal.pone.0251453>
- Amalina, I. K., & Vidákovich, T. (2023a). Development and differences in mathematical problem-solving skills: A cross-sectional study of differences in demographic backgrounds. *Heliyon, 9*(5), e16366. <https://doi.org/10.1016/j.heliyon.2023.e16366>
- Amalina, I. K., & Vidákovich, T. (2023b). Cognitive and socioeconomic factors that influence the mathematical problem-solving skills of students. *Heliyon, 9*(9), e19539. <https://doi.org/10.1016/j.heliyon.2023.e19539>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Ashcraft, M. H., Krause, J. A., & Hopko, D. R. (2007). *Is math anxiety a mathematical learning disability? Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (D. B. Berc). Paul H. Brookes Publishing Co.
- Baars, M., Wijnia, L., & Paas, F. (2017). The Association between Motivation, Affect, and Self-regulated Learning When Solving Problems. *Frontiers in Psychology, 8*. <https://doi.org/10.3389/fpsyg.2017.01346>
- Baker, E. ., & Mayer, R. . (1999). Computer-based assessment of problem solving. *Computers in Human Behavior, 15*(3–4), 269–282. [https://doi.org/10.1016/S0747-5632\(99\)00023-0](https://doi.org/10.1016/S0747-5632(99)00023-0)
- Bankel, J., Berggren, K.-F., Blom, K., Crawley, E. F., Wiklund, I., & Östlund, S. (2003). The CDIO syllabus: a comparative study of expected student proficiency. *European Journal of Engineering Education, 28*(3), 297–315. <https://doi.org/10.1080/0304379031000098274>
- Baron, P., & Corbin, L. (2012). Student engagement: rhetoric and reality. *Higher Education Research & Development, 31*(6), 759–772. <https://doi.org/10.1080/07294360.2012.655711>

- Bartley, J. E., Boeving, E. R., Riedel, M. C., Bottenhorn, K. L., Salo, T., Eickhoff, S. B., Brewe, E., Sutherland, M. T., & Laird, A. R. (2018). Meta-analytic evidence for a core problem solving network across multiple representational domains. *Neuroscience & Biobehavioral Reviews*, *92*, 318–337. <https://doi.org/10.1016/j.neubiorev.2018.06.009>
- Bassili, J. N., & Brown, R. D. (2005). Implicit and Explicit Attitudes: Research, Challenges, and Theory. In *The handbook of attitudes* (In D. Alba, pp. 543–574). Lawrence Erlbaum Associates Publishers.
- Bat Or, M. (2014). FEATS Problem-Solving Scale in PPAT of children aged 5–6.5 as related to their executive functions and motivation. *The Arts in Psychotherapy*, *41*(1), 27–35. <https://doi.org/10.1016/j.aip.2013.11.008>
- Bátor, J., & Szeberényi, J. (2021). Problem solving in the time of coronavirus pandemic. *Biochemistry and Molecular Biology Education*, *49*(6), 882–887. <https://doi.org/10.1002/bmb.21571>
- Bećirović, S. (2017). The Relationship between Gender, Motivation and Achievement in Learning English as a Foreign Language. *European Journal of Contemporary Education*, *6*(2), 210–220. <https://doi.org/10.13187/ejced.2017.2.210>
- Beckmann, J., & Goode, N. (2017). Missing the Wood for the Wrong Trees: On the Difficulty of Defining the Complexity of Complex Problem Solving Scenarios. *Journal of Intelligence*, *5*(2), 15. <https://doi.org/10.3390/jintelligence5020015>
- Bell, E. C., Willson, M. C., Wilman, A. H., Dave, S., & Silverstone, P. H. (2006). Males and females differ in brain activation during cognitive tasks. *NeuroImage*, *30*(2), 529–538. <https://doi.org/10.1016/j.neuroimage.2005.09.049>
- Bereby-Meyer, Y., & Kaplan, A. (2005). Motivational influences on transfer of problem-solving strategies. *Contemporary Educational Psychology*, *30*(1), 1–22. <https://doi.org/10.1016/j.cedpsych.2004.06.003>
- Blum, D., Holling, H., Galibert, M. S., & Forthmann, B. (2016). Task difficulty prediction of figural analogies. *Intelligence*, *56*, 72–81. <https://doi.org/10.1016/j.intell.2016.03.001>
- Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch Model* (Psychology Press (ed.); 2nd Editio). Psychology Press. <https://doi.org/10.4324/9781410614575>
- Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE—Life Sciences Education*, *15*(4), rm4. <https://doi.org/10.1187/cbe.16-04-0148>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer Science & Business Media.
- Borgonovi, F., & Greiff, S. (2020). Societal level gender inequalities amplify gender gaps in problem solving more than in academic disciplines. *Intelligence*, *79*, 101422. <https://doi.org/10.1016/j.intell.2019.101422>
- Brossard, D., Lewenstein, B., & Bonney, R. (2005). Scientific knowledge and attitude change: The impact of a citizen science project. *International Journal of Science Education*, *27*(9), 1099–1121. <https://doi.org/10.1080/09500690500069483>
- Brunner, M., Krauss, S., & Kunter, M. (2008). Gender differences in mathematics: Does the story need to be rewritten? *Intelligence*, *36*(5), 403–421. <https://doi.org/10.1016/j.intell.2007.11.002>

- BSNP. (2013). *Standar isi: standar kompetensi dan kompetensi dasar*.  
<https://kurikulum.kemdikbud.go.id/standar-nasional-pendidikan/>
- Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: An assessment of scientific literacy. *Journal of Research in Science Teaching*, 46(8), 865–883.  
<https://doi.org/10.1002/tea.20333>
- Cabras, C., Konyukhova, T., Lukianova, N., Mondo, M., & Sechi, C. (2023). Gender and country differences in academic motivation, coping strategies, and academic burnout in a sample of Italian and Russian first-year university students. *Heliyon*, 9(6), e16617.  
<https://doi.org/10.1016/j.heliyon.2023.e16617>
- Campos-Sánchez, A., López-Núñez, J. A., Carriel, V., Martín-Piedra, M.-Á., Sola, T., & Alaminos, M. (2014). Motivational component profiles in university students learning histology: a comparative study between genders and different health science curricula. *BMC Medical Education*, 14(1), 46. <https://doi.org/10.1186/1472-6920-14-46>
- Care, E., Scoular, C., & Griffin, P. (2016). Assessment of Collaborative Problem Solving in Education Environments. *Applied Measurement in Education*, 29(4), 250–264.  
<https://doi.org/10.1080/08957347.2016.1209204>
- 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>
- Carreira, J. M. (2011). Relationship between motivation for learning EFL and intrinsic motivation for learning in general among Japanese elementary school students. *System*, 39(1), 90–102. <https://doi.org/10.1016/j.system.2011.01.009>
- Cassidy, T., & Giles, M. (2009). Achievement Motivation, Problem-solving style, and Performance in Higher Education. *The Irish Journal of Psychology*, 30(3–4), 211–222.  
<https://doi.org/10.1080/03033910.2009.10446311>
- Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, 140(4), 980–1008. <https://doi.org/10.1037/a0035661>
- Chang, C.-Y. (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>
- Chang, Y., Chang, C., & Tseng, Y. (2010). Trends of science education research: an automatic content analysis. *Journal of Science Education and Technology*, 19(4), 315–331. <https://doi.org/10.1007/s10956-009-9202-2>
- Chao, C.-Y., Chen, Y.-T., & Chuang, K.-Y. (2015). Exploring students' learning attitude and achievement in flipped learning supported computer aided design curriculum: A study in high school engineering education. *Computer Applications in Engineering Education*, 23(4), 514–526. <https://doi.org/10.1002/cae.21622>
- Cheng, L.-T., Smith, T. J., Hong, Z.-R., & Lin, H. (2021). Gender and STEM background as predictors of college students' competencies in forming research questions and designing experiments in inquiry activities. *International Journal of Science Education*, 43(17), 2866–2883. <https://doi.org/10.1080/09500693.2021.1994167>
- Cheng, S.-C., She, H.-C., & Huang, L.-Y. (2017). The Impact of Problem-Solving Instruction

- on Middle School Students' Physical Science Learning: Interplays of Knowledge, Reasoning, and Problem Solving. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(3). <https://doi.org/10.12973/ejmste/80902>
- Chiang, P.-M., & Tzou, H.-I. (2018). The application of differential person functioning on the science literacy of Taiwan PISA 2015. *Humanities & Social Sciences Reviews*, 6(1), 08–13. <https://doi.org/10.18510/hssr.2018.612>
- Chiu, M.-S. (2022). Transcend socioeconomic status constraints to mathematics and science achievement by collaborative problem-solving: The female people-smartness hypothesis. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.944329>
- Christ, A., Becker, N., & Kröner, S. (2020). Multiple complex problem-solving scenarios: The incremental validity of ability self-concept beyond reasoning in adults. *Intelligence*, 78, 101421. <https://doi.org/10.1016/j.intell.2019.101421>
- Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, 38(2), 383–397. <https://doi.org/10.1111/cogs.12099>
- Christou, C., & Papageorgiou, E. (2007). A framework of mathematics inductive reasoning. *Learning and Instruction*, 17(1), 55–66. <https://doi.org/10.1016/j.learninstruc.2006.11.009>
- Chung, L.-Y., & Chang, R.-C. (2017). The Effect of Gender on Motivation and Student Achievement in Digital Game-based Learning: A Case Study of a Contented-Based Classroom. *EURASIA Journal of Mathematics, Science and Technology Education*, 13(6). <https://doi.org/10.12973/eurasia.2017.01227a>
- Crawley, F. E., & Koballa, T. R. (1994). Attitude research in science education: Contemporary models and methods. *Science Education*, 78(1), 35–55. <https://doi.org/10.1002/sce.3730780103>
- Csapó, B. (1997). The development of inductive reasoning: Cross-sectional assessments in an educational context. *International Journal of Behavioral Development*, 20(4), 609–626. <https://www.tandfonline.com/doi/abs/10.1080/016502597385081>
- Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: the eDia system. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01522>
- Csapó, Benő, & Funke, J. (2017). *The Nature of Problem Solving* (Benő Csapó & J. Funke (eds.)). OECD. <https://doi.org/10.1787/9789264273955-en>
- Daher, W., Alfahel, E., & Anabousy, A. (2021). Moderating the Relationship Between Student's Gender and Science Motivation. *Eurasia Journal of Mathematics, Science and Technology Education*, 17(5), em1956. <https://doi.org/10.29333/ejmste/10829>
- Davidson, J. E., Sternberg, R. J., & Sternberg, R. J. (2003). *The psychology of problem solving*. Cambridge University Press.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and Intrinsic Motivation to Use Computers in the Workplace1. *Journal of Applied Social Psychology*, 22(14), 1111–1132. <https://doi.org/10.1111/j.1559-1816.1992.tb00945.x>
- de Koning, E., Sijtsma, K., & Hamers, J. H. M. (2003). Construction and validation of test for inductive reasoning. *European Journal of Psychological Assessment*, 9(1), 24.

<https://doi.org/https://doi.org/10.1027/1015-5759.19.1.24>

- Debacker, T. K., & Nelson, R. M. (2000). Motivation to Learn Science: Differences Related to Gender, Class Type, and Ability. *The Journal of Educational Research*, 93(4), 245–254. <https://doi.org/10.1080/00220670009598713>
- Dermitzaki, I., Leondari, A., & Goudas, M. (2009). Relations between young students' strategic behaviours, domain-specific self-concept, and performance in a problem-solving situation. *Learning and Instruction*, 19(2), 144–157. <https://doi.org/10.1016/j.learninstruc.2008.03.002>
- Dermitzaki, I., Stavroussi, P., Vavougiou, D., & Kotsis, K. T. (2013). Adaptation of the Students' Motivation Towards Science Learning (SMTSL) questionnaire in the Greek language. *European Journal of Psychology of Education*, 28(3), 747–766. <https://doi.org/10.1007/s10212-012-0138-1>
- Díaz-Morales, J. F., & Escribano, C. (2013). Predicting school achievement: The role of inductive reasoning, sleep length and morningness–eveningness. *Personality and Individual Differences*, 55(2), 106–111. <https://doi.org/10.1016/j.paid.2013.02.011>
- Dindar, M. (2018). An empirical study on gender, video game play, academic success and complex problem solving skills. *Computers & Education*, 125, 39–52. <https://doi.org/10.1016/j.compedu.2018.05.018>
- Dindar, M., Järvelä, S., Nguyen, A., Haataja, E., & Çini, A. (2022). Detecting shared physiological arousal events in collaborative problem solving. *Contemporary Educational Psychology*, 69, 102050. <https://doi.org/10.1016/j.cedpsych.2022.102050>
- Ding, N., Bosker, R. J., & Harskamp, E. G. (2011). Exploring gender and gender pairing in the knowledge elaboration processes of students using computer-supported collaborative learning. *Computers & Education*, 56(2), 325–336. <https://doi.org/10.1016/j.compedu.2010.06.004>
- Dismore, H., & Bailey, R. (2011). Fun and enjoyment in physical education: young people's attitudes. *Research Papers in Education*, 26(4), 499–516. <https://doi.org/10.1080/02671522.2010.484866>
- Driscoll, M. M., & Burner, K. J. (2005). *Psychology of learning for instruction* (Third Edit). Pearson Allyn and Bacon.
- Dunbar, K. (2000). How Scientists Think in the Real World. *Journal of Applied Developmental Psychology*, 21(1), 49–58. [https://doi.org/10.1016/S0193-3973\(99\)00050-7](https://doi.org/10.1016/S0193-3973(99)00050-7)
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 61, 101859. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13, 378–385. <https://doi.org/https://doi.org/10.3758/BF03193858>
- Evans, J. S. B. T., Over, D. E., & Handley, S. J. (2003). Thinking: psychological perspectives on reasoning, judgment and decision making. In *A theory of hypothetical thinking* (pp. 3–22). John Wiley & Sons.

- Fang, Z., & Wei, Y. (2010). Improving Middle School Students' Science Literacy Through Reading Infusion. *The Journal of Educational Research*, 103(4), 262–273. <https://doi.org/10.1080/00220670903383051>
- Fawcett, L. M., & Garton, A. F. (2005). The effect of peer collaboration on children's problem-solving ability. *British Journal of Educational Psychology*, 75(2), 157–169. <https://doi.org/10.1348/000709904X23411>
- Ferreira, M. M., & Trudel, A. R. (2012). The impact of problem-based learning (PBL) on student attitude toward science, problem-solving skills, and sense of community in the classroom. *The Journal of Classroom Interaction*, 41(1), 23–30. <https://www.jstor.org/stable/43858871>
- Ferreira, M., & Trudel, A. (2012). The Impact of Problem-Based Learning (PBL) on Student Attitudes Toward Science, Problem-Solving Skills, and Sense of Community in the Classroom. *The Journal of Classroom Interaction*, 47(1), 23–30.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (M. Carmichael (ed.); 4th ed.). SAGE Publications Ltd.
- Fischer, A., Greiff, S., & Funke, J. (2012). The Process of Solving Complex Problems. *Journal of Problem Solving*, 4(1), 19–42. <https://ssrn.com/abstract=2097374>
- Fortus, D., Krajcik, J., Dershimer, R. C., Marx, R. W., & Mamlok-Naaman, R. (2005). Design-based science and real-world problem-solving. *International Journal of Science Education*, 27(7), 855–879. <https://doi.org/10.1080/09500690500038165>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Friege, G., & Lind, G. (2006). Types and Qualities of Knowledge and their Relations to Problem Solving in Physics. *International Journal of Science and Mathematics Education*, 4(3), 437–465. <https://doi.org/10.1007/s10763-005-9013-8>
- Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning*, 7(1), 69–89. <https://doi.org/10.1080/13546780042000046>
- Funke, J. (2010). Complex problem solving: a case for complex cognition? *Cognitive Processing*, 11(2), 133–142. <https://doi.org/10.1007/s10339-009-0345-0>
- Furinghetti, F., & Morselli, F. (2009). Every unsuccessful problem solver is unsuccessful in his or her own way: affective and cognitive factors in proving. *Educational Studies in Mathematics*, 70(1), 71–90. <https://doi.org/10.1007/s10649-008-9134-4>
- Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender Differences in Advanced Mathematical Problem Solving. *Journal of Experimental Child Psychology*, 75(3), 165–190. <https://doi.org/10.1006/jecp.1999.2532>
- Ganley, C. M., Vasilyeva, M., & Dulaney, A. (2014). Spatial Ability Mediates the Gender Difference in Middle School Students' Science Performance. *Child Development*, 85(4), 1419–1432. <https://doi.org/10.1111/cdev.12230>
- Gasco Txabarri, J., & Villarroel Villamor, J. D. (2017). La Motivación en la Resolución de Problemas Aritmético-algebraicos. Un Estudio con Alumnado de Educación Secundaria.

*Electronic Journal of Research in Education Psychology*, 12(32), 83–106.  
<https://doi.org/10.14204/ejrep.32.13076>

- Gauld, C. (1982). The scientific attitude and science education: A critical reappraisal. *Science Education*, 66(1), 109–121.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132(5), 692–731. <https://doi.org/10.1037/0033-2909.132.5.692>
- Glaser, R., Chi, M. T. H., & Farr, M. J. (2009). *Measuring complex problem solving: The MicroDYN approach*.
- Gliner, J. A., Morgan, G. A., & Leech, N. L. (2016). *Research methods in applied settings: An integrated approach to design and analysis*. Routledge.
- Glock, S., & Kovacs, C. (2013). Educational Psychology: Using Insights from Implicit Attitude Measures. *Educational Psychology Review*, 25(4), 503–522.  
<https://doi.org/10.1007/s10648-013-9241-3>
- Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, 48(10), 1159–1176. <https://doi.org/10.1002/tea.20442>
- Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2009). Science Motivation Questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching*, 46(2), 127–146. <https://doi.org/10.1002/tea.20267>
- Gok, T., & Silay, I. (2010). The Effects of Problem Solving Strategies on Students' Achievement, Attitude and Motivation. *Latin-American Journal of Physics Education*, 4(1), 7–21. [http://www.lajpe.org/jan10/02\\_Tolga\\_Gok.pdf](http://www.lajpe.org/jan10/02_Tolga_Gok.pdf)
- Gok, T. (2014). Peer instruction in the physics classroom: Effects on gender difference performance, conceptual learning, and problem solving. *Journal of Baltic Science Education*, 13(6).
- Gok, Tolga. (2014). Peer Instruction in The Physics Classroom: Effects on Gender Difference Performance, Conceptual Learning, and Problem Solving. *Journal of Baltic Science Education*, 13(6), 776–788. <https://doi.org/10.33225/jbse/14.13.776>
- Goode, N., & Beckmann, J. F. (2010). You need to know: There is a causal relationship between structural knowledge and control performance in complex problem solving tasks. *Intelligence*, 38(3), 345–352. <https://doi.org/10.1016/j.intell.2010.01.001>
- Granello, D. H. (2001). Promoting Cognitive Complexity in Graduate Written Work: Using Bloom's Taxonomy as a Pedagogical Tool to Improve Literature Reviews. *Counselor Education and Supervision*, 40(4), 292–307. <https://doi.org/10.1002/j.1556-6978.2001.tb01261.x>
- Greiff, S. (2012). Assessment and Theory in Complex Problem Solving - A Continuing Contradiction? *Journal of Educational and Developmental Psychology*, 2(1).  
<https://doi.org/10.5539/jedp.v2n1p49>
- Greiff, S. ., & Funke, J. (2017). Interactive problem solving: Exploring the potential of minimal complex systems. In *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning* (Benő Csapó, pp. 93–105). OECD Publishing.

<https://doi.org/10.1787/9789264273955-8-en>

- Greiff, S., Fischer, A., Stadler, M., & Wüstenberg, S. (2015). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning, 21*(3), 356–382. <https://doi.org/10.1080/13546783.2014.989263>
- Greiff, S., Molnár, G., Martin, R., Zimmermann, J., & Csapó, B. (2018). Students' exploration strategies in computer-simulated complex problem environments: A latent class approach. *Computers & Education, 126*, 248–263. <https://doi.org/10.1016/j.compedu.2018.07.013>
- Greiff, S., & Neubert, J. C. (2014). On the relation of complex problem solving, personality, fluid intelligence, and academic achievement. *Learning and Individual Differences, 36*, 37–48. <https://doi.org/10.1016/j.lindif.2014.08.003>
- Greiff, S., Stadler, M., Sonnleitner, P., Wolff, C., & Martin, R. (2015). Sometimes less is more: Comparing the validity of complex problem solving measures. *Intelligence, 50*, 100–113. <https://doi.org/10.1016/j.intell.2015.02.007>
- Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review, 13*, 74–83. <https://doi.org/10.1016/j.edurev.2014.10.002>
- Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic Problem Solving. *Applied Psychological Measurement, 36*(3), 189–213. <https://doi.org/10.1177/0146621612439620>
- Gunawan, G., Mashami, R. A., & Herayanti, L. (2020). Gender Description on Problem-Solving Skills in Chemistry Learning Using Interactive Multimedia. *Journal for the Education of Gifted Young Scientists, 8*(1), 571–589. <https://doi.org/10.17478/jegys.627095>
- Guven, B., & Cabakcor, B. O. (2013). Factors influencing mathematical problem-solving achievement of seventh grade Turkish students. *Learning and Individual Differences, 23*, 131–137. <https://doi.org/10.1016/j.lindif.2012.10.003>
- Hacieminoglu, E. (2016). Elementary School Students' Attitude toward Science and Related Variables. *The International Journal of Environmental and Science Education, 11*(2), 35–52. <https://doi.org/10.12973/ijese.2016.288a>
- Halpern, D. F., & LaMay, M. L. (2000). The Smarter Sex: A Critical Review of Sex Differences in Intelligence. *Educational Psychology Review, 12*(2), 229–246. <https://doi.org/10.1023/A:1009027516424>
- Hambrick, D. Z., & Engle, R. . (2003). The role of working memory in problem solving. In *The psychology of problem solving* (pp. 176–206). Cambridge University Pres.
- Hamers, J. H. M., de Koning, E., & Sijtsma, K. (1998). Inductive Reasoning in Third Grade: Intervention Promises and Constraints. *Contemporary Educational Psychology, 23*(2), 132–148. <https://doi.org/10.1006/ceps.1998.0966>
- Harskamp, E., Ding, N., & Suhre, C. (2008). Group composition and its effect on female and male problem-solving in science education. *Educational Research, 50*(4), 307–318. <https://doi.org/10.1080/00131880802499688>

- Haverty, L. A., Koedinger, K. R., Klahr, D., & Alibali, M. W. (2000). Solving Inductive Reasoning Problems in Mathematics: Not-so-Trivial Pursuit. *Cognitive Science*, 24(2), 249–298. [https://doi.org/10.1207/s15516709cog2402\\_3](https://doi.org/10.1207/s15516709cog2402_3)
- Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *WIREs Cognitive Science*, 1(2), 278–292. <https://doi.org/10.1002/wcs.44>
- Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence*, 48, 1–14. <https://doi.org/https://doi.org/10.1016/j.intell.2014.10.005>
- He, W., & Wong, W. (2021). Gender Differences in the Distribution of Creativity Scores: Domain-Specific Patterns in Divergent Thinking and Creative Problem Solving. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.626911>
- Herrmann, W., Beckmann, J. F., & Kretzschmar, A. (2023). The role of learning in complex problem solving using MicroDYN. *Intelligence*, 100, 101773. <https://doi.org/10.1016/j.intell.2023.101773>
- Hoegh-Guldberg, O. (2011). The Impact of Climate Change on Coral Reef Ecosystems. In *Coral Reefs: An Ecosystem in Transition* (pp. 391–403). Springer Netherlands. [https://doi.org/10.1007/978-94-007-0114-4\\_22](https://doi.org/10.1007/978-94-007-0114-4_22)
- Hoffman, B., & Schraw, G. (2009). The influence of self-efficacy and working memory capacity on problem-solving efficiency. *Learning and Individual Differences*, 19(1), 91–100. <https://doi.org/10.1016/j.lindif.2008.08.001>
- Holyoak, K. J. (2012). Analogy and relational reasoning. In *The Oxford handbook of thinking and reasoning* (pp. 234–259). Oxford University Press.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, P.-S., & Chen, H.-C. (2016). Gender Differences in Eye Movements in Solving Text-and-Diagram Science Problems. *International Journal of Science and Mathematics Education*, 14(S2), 327–346. <https://doi.org/10.1007/s10763-015-9644-3>
- Hung, C. M., Hwang, G. J., & Huang, I. (2012). A project-based digital storytelling approach for improving students' learning motivation, problem-solving competence and learning achievement. *Journal of Educational Technology & Society*, 15(4), 368–379.
- Hyde, J. S. (2016). Sex and cognition: gender and cognitive functions. *Current Opinion in Neurobiology*, 38, 53–56. <https://doi.org/10.1016/j.conb.2016.02.007>
- Ichien, N., Lu, H., & Holyoak, K. J. (2020). Verbal analogy problem sets: An inventory of testing materials. *Behavior Research Methods*, 52(5), 1803–1816. <https://doi.org/10.3758/s13428-019-01312-3>
- Ifenthaler, D., & Seel, N. M. (2011). A longitudinal perspective on inductive reasoning tasks. Illuminating the probability of change. *Learning and Instruction*, 21(4), 538–549. <https://doi.org/10.1016/j.learninstruc.2010.08.004>
- Japelj Pavešić, B. (2008). Science achievement, gender differences, and experimental work in classes in Slovenia as evident in TIMSS studies. *Studies in Educational Evaluation*, 34(2), 94–104. <https://doi.org/10.1016/j.stueduc.2008.04.005>

- Johnson-Laird, P. N., & Khemlani, S. S. (2013). Toward a Unified Theory of Reasoning. In *Psychology of Learning and Motivation* (59th ed., pp. 1–46). Elsevier Ltd. <https://doi.org/https://doi.org/10.1016/B978-0-12-407187-2.00001-0>
- Jonassen, D. H. (1997). Instructional design models for well-structured and III-structured problem-solving learning outcomes. *Educational Technology Research and Development*, 45(1), 65–94. <https://doi.org/10.1007/BF02299613>
- Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology*, 96(5), 933–948. <https://doi.org/10.1037/a0014747>
- Jones, M. G., Howe, A., & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science Education*, 84(2), 180–192. [https://doi.org/10.1002/\(SICI\)1098-237X\(200003\)84:2<180::AID-SCE3>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1098-237X(200003)84:2<180::AID-SCE3>3.0.CO;2-X)
- Kambeyo, L. (2018). *Assessing Namibian students' abilities in scientific reasoning, scientific inquiry and inductive reasoning skills*. University of Szeged.
- Kenny, M. E., Blustein, D. L., Haase, R. F., Jackson, J., & Perry, J. C. (2006). Setting the stage: Career development and the student engagement process. *Journal of Counseling Psychology*, 53(2), 272–279. <https://doi.org/10.1037/0022-0167.53.2.272>
- Kerlinger, F. N. (1970). *Foundations of behavioral research*. Rinehart and Winston Inc.
- Khine, M. S. (2013). *Application of Structural Equation Modeling in Educational Research and Practice* (Myint Swe Khine (ed.)). SensePublishers. <https://doi.org/10.1007/978-94-6209-332-4>
- Khine, M. S. (2015). *Attitude measurements in science education: Classic and contemporary approaches*. Information Age Publishing, INC.
- Khishfe, R., & BouJaoude, S. (2016). Lebanese Students' Conceptions of and Attitudes Towards Science and Related Careers Based on Their Gender and Religious Affiliations. *International Journal of Science and Mathematics Education*, 14(S1), 145–167. <https://doi.org/10.1007/s10763-014-9587-0>
- Kim, C., & Pekrun, R. (2014). Emotions and Motivation in Learning and Performance. In *Handbook of Research on Educational Communications and Technology* (Vol. 58, Issue 5, pp. 65–75). Springer New York. [https://doi.org/10.1007/978-1-4614-3185-5\\_6](https://doi.org/10.1007/978-1-4614-3185-5_6)
- Kim, J. H., Park, S., Kim, S. H., & Lee, E. J. (2021). Identifying high-priority conservation areas for endangered waterbirds using a flagship species in the Korean DMZ. *Ecological Engineering*, 159, 106080. <https://doi.org/10.1016/j.ecoleng.2020.106080>
- Kind, P., Jones, K., & Barmby, P. (2007). Developing Attitudes towards Science Measures. *International Journal of Science Education*, 29(7), 871–893. <https://doi.org/10.1080/09500690600909091>
- Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: a training approach. *Review of Educational Research*, 78(1), 85–123. <https://doi.org/10.3102/0034654307313402>
- Klauer, Karl. Josef., Willmes, K., & Phye, G. D. (2002). Inducing Inductive Reasoning: Does It Transfer to Fluid Intelligence? *Contemporary Educational Psychology*, 27(1), 1–25. <https://doi.org/10.1006/ceps.2001.1079>
- Köksal, D., & Ulum, Ö. G. (2018). Language assessment through Bloom's Taxonomy.

- Journal of Language and Linguistic Studies*, 14(2), 76–88.  
<https://www.jlls.org/index.php/jlls/article/view/841/372>
- Korur, F., Vargas, R., & Serrano, N. T. (2016). Attitude toward Science Teaching of Spanish and Turkish In-service Elementary Teachers: Multi-group Confirmatory Factor Analysis. *EURASIA Journal of Mathematics, Science and Technology Education*, 12(2).  
<https://doi.org/10.12973/eurasia.2016.1215a>
- Krogh, L. B., & Thomsen, P. V. (2005). Studying students' attitudes towards science from a cultural perspective but with a quantitative methodology: border crossing into the physics classroom. *International Journal of Science Education*, 27(3), 281–302.  
<https://doi.org/10.1080/09500690412331314469>
- Kroner, S., Plass, J., & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence*, 33(4), 347–368. <https://doi.org/10.1016/j.intell.2005.03.002>
- Kunda, M., McGregor, K., & Goel, A. K. (2013). A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations. *Cognitive Systems Research*, 22, 47–66.  
<https://doi.org/https://doi.org/10.1016/j.cogsys.2012.08.001>
- Kuo, F.-R., Hwang, G.-J., & Lee, C.-C. (2012). A hybrid approach to promoting students' web-based problem-solving competence and learning attitude. *Computers & Education*, 58(1), 351–364. <https://doi.org/10.1016/j.compedu.2011.09.020>
- Kusurkar, R. A., Ten Cate, T. J., Vos, C. M. P., Westers, P., & Croiset, G. (2013). How motivation affects academic performance: a structural equation modelling analysis. *Advances in Health Sciences Education*, 18(1), 57–69. <https://doi.org/10.1007/s10459-012-9354-3>
- Kyllonen, P., & Kell, H. (2017). Fluid Ability (Gf) and Complex Problem Solving (CPS). *Journal of Intelligence*, 5(3), 28. <https://doi.org/10.3390/jintelligence5030028>
- Lachman, R., Lachman, J. L., & Butterfield, E. C. (2015). *Cognitive psychology and information processing: An introduction*. Psychology Press.
- Lai, Y., Zhu, X., Chen, Y., & Li, Y. (2015). Effects of Mathematics Anxiety and Mathematical Metacognition on Word Problem Solving in Children with and without Mathematical Learning Difficulties. *PLOS ONE*, 10(6), e0130570.  
<https://doi.org/10.1371/journal.pone.0130570>
- Lawson, A. E., Banks, D. L., & Logvin, M. (2007). Self-efficacy, reasoning ability, and achievement in college biology. *Journal of Research in Science Teaching*, 44(5), 706–724. <https://doi.org/10.1002/tea.20172>
- Lee-Cultura, S., Sharma, K., & Giannakos, M. (2022). Children's play and problem-solving in motion-based learning technologies using a multi-modal mixed methods approach. *International Journal of Child-Computer Interaction*, 31, 100355.  
<https://doi.org/10.1016/j.ijcci.2021.100355>
- Leighton, J. P., & Sternberg, R. J. (2004). *The Nature of Reasoning*. Cambridge University Press.
- Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and Extrinsic Motivational Orientations in the Classroom: Age Differences and Academic Correlates. *Journal of Educational Psychology*, 97(2), 184–196. <https://doi.org/10.1037/0022-0663.97.2.184>

- Li, L., Zhou, X., Huang, J., Tu, D., Gao, X., Yang, Z., & Li, M. (2020). Assessing kindergarteners' mathematics problem solving: The development of a cognitive diagnostic test. *Studies in Educational Evaluation*, *66*, 100879. <https://doi.org/10.1016/j.stueduc.2020.100879>
- Liang, P., Jia, X., Taatgen, N. A., Zhong, N., & Li, K. (2014). Different strategies in solving series completion inductive reasoning problems: An fMRI and computational study. *International Journal of Psychophysiology*, *93*(2), 253–260. <https://doi.org/10.1016/j.ijpsycho.2014.05.006>
- Lin, H., Hong, Z.-R., & Huang, T.-C. (2012). The Role of Emotional Factors in Building Public Scientific Literacy and Engagement with Science. *International Journal of Science Education*, *34*(1), 25–42. <https://doi.org/10.1080/09500693.2010.551430>
- Liu, M., Hu, W., Jiannong, S., & Adey, P. (2010). Gender Stereotyping and Affective Attitudes Towards Science in Chinese Secondary School Students. *International Journal of Science Education*, *32*(3), 379–395. <https://doi.org/10.1080/09500690802595847>
- Liu, Ting-ting, & Sun, H. (2012). Gender Differences on Information Literacy of Science and Engineering Undergraduates. *International Journal of Modern Education and Computer Science*, *4*(2), 23–30. <https://doi.org/10.5815/ijmecs.2012.02.04>
- Liu, Tongxi, & Israel, M. (2022). Uncovering students' problem-solving processes in game-based learning environments. *Computers & Education*, *182*, 104462. <https://doi.org/10.1016/j.compedu.2022.104462>
- Lotz, C., Scherer, R., Greiff, S., & Sparfeldt, J. R. (2022). g's little helpers – VOTAT and NOTAT mediate the relation between intelligence and complex problem solving. *Intelligence*, *95*, 101685. <https://doi.org/10.1016/j.intell.2022.101685>
- Lotz, C., Sparfeldt, J. R., & Greiff, S. (2016). Complex problem solving in educational contexts – Still something beyond a “good g”? *Intelligence*, *59*, 127–138. <https://doi.org/10.1016/j.intell.2016.09.001>
- Luo, R.-Z., Zhang, X.-H., Zhang, C.-M., & Liu, Y.-H. (2019). Impact of self-directed learning readiness and learning attitude on problem-solving ability among Chinese undergraduate nursing students. *Frontiers of Nursing*, *6*(2), 143–150. <https://doi.org/10.2478/FON-2019-0021>
- Luu, K., & Freeman, J. G. (2011). An analysis of the relationship between information and communication technology (ICT) and scientific literacy in Canada and Australia. *Computers & Education*, *56*(4), 1072–1082. <https://doi.org/10.1016/j.compedu.2010.11.008>
- Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: a longitudinal panel analysis. *Journal of Adolescence*, *27*(2), 165–179. <https://doi.org/10.1016/j.adolescence.2003.11.003>
- MacIntyre, P. D., Baker, S. C., Clément, R., & Donovan, L. A. (2002). Sex and Age Effects on Willingness to Communicate, Anxiety, Perceived Competence, and L2 Motivation Among Junior High School French Immersion Students. *Language Learning*, *52*(3), 537–564. <https://doi.org/10.1111/1467-9922.00194>
- Magiera, M. (2012). K-8 Preservice Teachers' Inductive Reasoning in the Problem-Solving Contexts. *Mathematics, Statistics and Computer Science Faculty Research and*

- Publications*, 84, 1–26. [https://epublications.marquette.edu/mscs\\_fac/84](https://epublications.marquette.edu/mscs_fac/84)
- Marchis, I. (2013). Relation between students' attitude towards mathematics and their problem solving skills. *PedActa*, 3(2), 59–66. <http://padi.psiedu.ubbcluj.ro/pedacta/>
- Mark, G. L., & Brooke, A. E. (2006). Tomato disease and disorder. In *Disease in outdoor production* (pp. 1–12). ISU Extension Publication.
- Martin-Dunlop, C., & Fraser, B. J. (2007). Learning Environment and Attitudes Associated with an Innovative Science Course Designed for Prospective Elementary Teachers. *International Journal of Science and Mathematics Education*, 6(1), 163–190. <https://doi.org/10.1007/s10763-007-9070-2>
- Martin, A. J. (2009). Age appropriateness and motivation, engagement, and performance in high school: Effects of age within cohort, grade retention, and delayed school entry. *Journal of Educational Psychology*, 101(1), 101–114. <https://doi.org/10.1037/a0013100>
- Martin, A. J., Liem, G. A. D., Mok, M. M. C., & Xu, J. (2012). Problem solving and immigrant student mathematics and science achievement: Multination findings from the Programme for International Student Assessment (PISA). *Journal of Educational Psychology*, 104(4), 1054–1073. <https://doi.org/10.1037/a0029152>
- Mayer, R. E., & Wittrock, M. C. (2006). Problem Solving. In Mahwah (Ed.), *Handbook of Educational Psychology* (PA Alexand, pp. 287–303). Erlbaum. [https://scholar.google.com/scholar?q=Mayer%2C+R.+E.%2C+%26+Wittrock%2C+M.+C.+%282006%29.+Problem+solving.+In+P.+A.+Alexander+%26+P.+H.+Winne+%28E+ds.%29%2C+Handbook+of+educational+psychology+%28pp.+287-303%29.+Mahwah%2C+NJ%3A+Erlbaum&hl=en&as\\_sdt=0%2C5&as\\_](https://scholar.google.com/scholar?q=Mayer%2C+R.+E.%2C+%26+Wittrock%2C+M.+C.+%282006%29.+Problem+solving.+In+P.+A.+Alexander+%26+P.+H.+Winne+%28E+ds.%29%2C+Handbook+of+educational+psychology+%28pp.+287-303%29.+Mahwah%2C+NJ%3A+Erlbaum&hl=en&as_sdt=0%2C5&as_)
- McAbee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review*, 27(2), 277–290. <https://doi.org/10.1016/j.hrmr.2016.08.005>
- Mercan, F. C. (2020). Control-value theory and enjoyment of science: A cross-national investigation with 15-year-olds using PISA 2006 data. *Learning and Individual Differences*, 80, 101889. <https://doi.org/10.1016/j.lindif.2020.101889>
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74. <https://doi.org/https://dx.doi.org/10.1017/S0140525X10000968>
- Mihladiz, G., Duran, M., & Dogan, A. (2011). Examining primary school students' attitudes towards science in terms of gender, class level and income level. *Procedia - Social and Behavioral Sciences*, 15, 2582–2588. <https://doi.org/10.1016/j.sbspro.2011.04.150>
- Molnár, G. (2011). Playful fostering of 6- to 8-year-old students' inductive reasoning. *Thinking Skills and Creativity*, 6(2), 91–99. <https://doi.org/10.1016/j.tsc.2011.05.002>
- 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>
- Molnár, Gyöngyvér, 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>

- Moore, A. M., Rudig, N. O., & Ashcraft, M. H. (2014). *Affect, Motivation, Working Memory, and Mathematics* (R. Cohen Kadosh & A. Dowker (eds.); Vol. 1). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199642342.013.004>
- Mousa, M., & Molnár, G. (2020). Computer-based training in math improves inductive reasoning of 9- to 11-year-old children. *Thinking Skills and Creativity*, 37, 100687. <https://doi.org/10.1016/j.tsc.2020.100687>
- Mueller, D. J. (1986). *Measuring social attitudes*. Teacher College Press.
- Mukhopadhyay, R. (2013). Problem solving in science learning- some important consideration of teacher. *IOSR Journal of Humanities and Social Science (IOSR-JHSS)*, 4(6), 21–25.
- Nadapdap, A. T. Y., & Istiyono, E. (2017). Developing physics problem-solving skill test for grade X students of senior high school. *Research and Evaluation in Education*, 3(2), 114. <https://doi.org/10.21831/reid.v3i2.14982>
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. <https://nap.nationalacademies.org/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts>
- Navarro, M., Förster, C., González, C., & González-Pose, P. (2016). Attitudes toward science: measurement and psychometric properties of the Test of Science-Related Attitudes for its use in Spanish-speaking classrooms. *International Journal of Science Education*, 38(9), 1459–1482. <https://doi.org/10.1080/09500693.2016.1195521>
- Nickles, T. (1988). Questioning and problems in philosophy of science: Problem-solving versus directly truth-seeking epistemologies. *Questions and Questioning*, 43–67.
- Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, 109(3), 504–511. <https://doi.org/10.1037/0021-843X.109.3.504>
- Nwosu, A. A., & Ibe, E. (2014). Gender and Scientific Literacy Levels: Implications for Sustainable Science and Technology Education (STE) for the 21st Century Jobs. *Journal of Education and Practice*, 5(8), 113–118. [https://doi.org/Gender and Scientific Literacy Levels: Implications for Sustainable Science and Technology Education \(STE\) for the 21st Century Jobs](https://doi.org/Gender and Scientific Literacy Levels: Implications for Sustainable Science and Technology Education (STE) for the 21st Century Jobs)
- OECD. (2004). *A Profile of Student Performance in Reading and Science*. OECD Publishing. <https://www.oecd-ilibrary.org/docserver/9789264006416-7-en.pdf?expires=1655722645&id=id&accname=guest&checksum=9D08792F6277FEC3D8E25CE40797CD6E>
- OECD. (2014). *PISA 2012 Results: Creative Problem Solving (Volume V)*. OECD. <https://doi.org/10.1787/9789264208070-en>
- OECD. (2016). *PISA 2015 Results (Volume I)*. OECD. <https://doi.org/10.1787/9789264266490-en>
- OECD. (2017a). *PISA 2015 Collaborative Problem-Solving Framework*. <https://www.oecd.org/pisa/pisaproducts/Draft PISA 2015 Collaborative Problem Solving Framework .pdf>

- OECD. (2017b). *PISA for development assessment and analytical framework: reading, mathematics and science, preliminary version*. OECD Publishing. <https://www.oecd.org/pisa/pisa-for-development/PISA-D-Assessment-and-Analytical-Framework-Ebook.pdf>
- OECD. (2017c). *PISA 2015 collaborative problem-solving framework* (pp. 131–188). OECD Publishing. <https://doi.org/10.1787/9789264281820-8-en>
- OECD. (2019a). *PISA 2018 Assessment and Analytical Framework*. OECD. <https://doi.org/10.1787/b25efab8-en>
- OECD. (2019b). *PISA 2018 Science Framework* (OECD Publi, pp. 97–117). <https://doi.org/10.1787/f30da688-en>
- OECD. (2019c). *PISA 2018 results (Volume I)*. OECD. <https://doi.org/10.1787/5f07c754-en>
- Olasehinde, K. J., & Olatoye, R. A. (2014). Scientific Attitude, Attitude to Science and Science Achievement of Senior Secondary School Students in Katsina State, Nigeria. *Journal of Educational and Social Research*. <https://doi.org/10.5901/jesr.2014.v4n1p445>
- Oon, P.-T., Cheng, M. M. W., & Wong, A. S. L. (2020). Gender differences in attitude towards science: methodology for prioritising contributing factors. *International Journal of Science Education*, 42(1), 89–112. <https://doi.org/10.1080/09500693.2019.1701217>
- Oppermann, E., Vinni-Laakso, J., Juuti, K., Loukomies, A., & Salmela-Aro, K. (2021). Elementary school students' motivational profiles across Finnish language, mathematics and science: Longitudinal trajectories, gender differences and STEM aspirations. *Contemporary Educational Psychology*, 64, 101927. <https://doi.org/10.1016/j.cedpsych.2020.101927>
- Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049–1079. <https://doi.org/10.1080/0950069032000032199>
- Oskamp, S., & Schultz, P. W. (2005). *Attitudes and Opinions*. Psychology Press. <https://doi.org/10.4324/9781410611963>
- Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2012). Anxiety and depression in academic performance: An exploration of the mediating factors of worry and working memory. *School Psychology International*, 33(4), 433–449. <https://doi.org/10.1177/0143034311427433>
- Ozbas, S. (2016). High School Students' Intrinsic and Extrinsic Regulation in Learning Biology. *International Journal of Educational Sciences*, 15(1–2), 261–268. <https://doi.org/10.1080/09751122.2016.11890535>
- Özcan, Z. Ç. (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>
- Öztürk, M., Akkan, Y., & Kaplan, A. (2020). Reading comprehension, Mathematics self-efficacy perception, and Mathematics attitude as correlates of students' non-routine Mathematics problem-solving skills in Turkey. *International Journal of Mathematical Education in Science and Technology*, 51(7), 1042–1058. <https://doi.org/10.1080/0020739X.2019.1648893>

- Ozturk, T., & Guven, B. (2016). Evaluating Students' Beliefs in Problem Solving Process: A Case Study. *EURASIA Journal of Mathematics, Science and Technology Education*, 12(3). <https://doi.org/10.12973/eurasia.2016.1208a>
- Park, J., & Lee, L. (2004). Analysing cognitive or non-cognitive factors involved in the process of physics problem-solving in an everyday context. *International Journal of Science Education*, 26(13), 1577–1595. <https://doi.org/10.1080/0950069042000230767>
- Pásztor-Kovács, A., Pásztor, A., & Molnár, G. (2023). Measuring collaborative problem solving: research agenda and assessment instrument. *Interactive Learning Environments*, 31(8), 5159–5179. <https://doi.org/10.1080/10494820.2021.1999273>
- Pásztor, A., Molnár, G., Korom, E. B., Németh, M., & Csapó, B. (2017). Online assessment of inductive reasoning and its predictive power on inquiry skills in science. *17th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI)*, 509.
- Pedaste, M., Mäeots, M., Siiman, L. A., Jong, T. De, Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning : Definitions and the inquiry cycle. *Educational Research Review*, 14, 47–61. <https://doi.org/10.1016/j.edurev.2015.02.003>
- Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Phillips, J. A., McCallum, J. E., Clemmer, K. W., & Zachariah, T. M. (2016). A Problem-Solving Framework to Assist Students and Teachers in STEM Courses. *Physics*. <https://doi.org/https://doi.org/10.48550/arXiv.1607.07853>
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40. <https://doi.org/10.1037/0022-0663.82.1.33>
- Pólya, G. (1945). *How to Solve It*. Princeton University Press.
- 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 & Technological Education*, 38(1), 105–130. <https://doi.org/10.1080/02635143.2019.1600490>
- Raccanello, D., Brondino, M., Moè, A., Stupnisky, R., & Lichtenfeld, S. (2019). Enjoyment, Boredom, Anxiety in Elementary Schools in Two Domains: Relations With Achievement. *The Journal of Experimental Education*, 87(3), 449–469. <https://doi.org/10.1080/00220973.2018.1448747>
- Rahmi, Y. L., Habibah, I. N., Zulyusri, Z., & Darussyamsu, R. (2021). HOTS assessment in circulatory system learning: Validity, reliability, and item quality. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 7(2), 171–178. <https://doi.org/10.22219/jpbi.v7i2.15513>
- Ralph, M. C., Schneider, B., Benson, D. R., & Ward, D. (2022). Separated by spaces: Undergraduate students re-sort along attitude divides when choosing whether to learn in spaces designed for active learning. *Active Learning in Higher Education*, 146978742211188. <https://doi.org/10.1177/14697874221118866>
- Ramírez-Uclés, I. M., & Ramírez-Uclés, R. (2020). Gender Differences in Visuospatial Abilities and Complex Mathematical Problem Solving. *Frontiers in Psychology*, 11.

<https://doi.org/10.3389/fpsyg.2020.00191>

- Rausch, A., & Wuttke, E. (2016). Development of multi faceted model of domain specific problem solving and its acceptance by different stakeholders in the business domain. *Unterrichtswissenschaft, 44*(2), 169–184.
- Reid, N. (2006). Thoughts on attitude measurement. *Research in Science & Technological Education, 24*(1), 3–27. <https://doi.org/10.1080/02635140500485332>
- Reilly, D. (2012). Gender, Culture, and Sex-Typed Cognitive Abilities. *PLoS ONE, 7*(7), e39904. <https://doi.org/10.1371/journal.pone.0039904>
- Reilly, D., & Neumann, D. L. (2013). Gender-Role Differences in Spatial Ability: A Meta-Analytic Review. *Sex Roles, 68*(9–10), 521–535. <https://doi.org/10.1007/s11199-013-0269-0>
- Reilly, D., Neumann, D. L., & Andrews, G. (2019). Investigating Gender Differences in Mathematics and Science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education, 49*(1), 25–50. <https://doi.org/10.1007/s11165-017-9630-6>
- Rittle-Johnson, B. (2017). Developing mathematics knowledge. *Child Development Perspectives, 11*(3), 184–190. <https://doi.org/10.1111/cdep.12229>
- Rivera, F., & Becker, J. R. (2003). The Effects of Numerical and Figural Cues on the Induction Processes of Preservice Elementary Teachers. *International Group for the Psychology of Mathematics Education, 4*, 63–70. <https://files.eric.ed.gov/fulltext/ED501090.pdf>
- Royer, J., & Garofoli, L. (2005). Cognitive contributions to sex differences in math performance. In *Gender differences in mathematics*. Cambridge University Press. <http://ndl.ethernet.edu.et/bitstream/123456789/16205/1/4.pdf#page=116>
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology, 25*(1), 54–67. <https://doi.org/10.1006/ceps.1999.1020>
- Salazar, M., & Salas, E. (2013). Reflections of Cross-Cultural Collaboration Science. *Journal of Organizational Behavior, 34*(6), 910–917. <https://doi.org/10.1002/job.1881>
- Salmi, H., & Thuneberg, H. (2019). The role of self-determination in informal and formal science learning contexts. *Learning Environments Research, 22*(1), 43–63. <https://doi.org/10.1007/s10984-018-9266-0>
- Sbeglia, G. C., & Nehm, R. H. (2019). Do you see what I-SEA? A Rasch analysis of the psychometric properties of the Inventory of Student Evolution Acceptance. *Science Education, 103*(2), 287–316. <https://doi.org/10.1002/sc.21494>
- Schaverien, L., & Cosgrove, M. (2000). A biological basis for generative learning in technology-and-science Part II: Implications for technology-and-science education. *International Journal of Science Education, 22*(1), 13–35. <https://doi.org/10.1080/095006900289985>
- Schoenfeld, A.H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast, 10*(1), 9–35. <https://scholarworks.umt.edu/tme/vol10/iss1/3>
- Schoenfeld, Alan H. (1987). Pólya, Problem Solving, and Education. *Mathematics Magazine,*

60(5), 283–291. <https://doi.org/10.1080/0025570X.1987.11977325>

- Schoenfeld, Alan H. (2016). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics (Reprint). *Journal of Education*, 196(2), 1–38. <https://doi.org/10.1177/002205741619600202>
- Schumm, M. F., & Bogner, F. X. (2016). Measuring adolescent science motivation. *International Journal of Science Education*, 38(3), 434–449. <https://doi.org/10.1080/09500693.2016.1147659>
- Schürmann, L., & Quaiser-Pohl, C. (2022). Out-of-school learning levels prior achievement and gender differences in secondary school students' motivation. *International Journal of Educational Research Open*, 3, 100158. <https://doi.org/10.1016/j.ijedro.2022.100158>
- Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences*, 24, 42–52. <https://doi.org/10.1016/j.lindif.2012.12.011>
- Seifried, J., Brandt, S., Kögler, K., & Rausch, A. (2020). The computer-based assessment of domain-specific problem-solving competence—A three-step scoring procedure. *Cogent Education*, 7(1). <https://doi.org/10.1080/2331186X.2020.1719571>
- Shahzadi, I., & Nasreen, A. (2020). Assessing Scientific Literacy Levels among Secondary School Science Students of District Lahore. *Bulletin of Education and Research*, 42(3), 1–21. <https://files.eric.ed.gov/fulltext/EJ1291080.pdf>
- She, H.-C., Cheng, M.-T., Li, T.-W., Wang, C.-Y., Chiu, H.-T., Lee, P.-Z., Chou, W.-C., & Chuang, M.-H. (2012). Web-based undergraduate chemistry problem-solving: The interplay of task performance, domain knowledge and web-searching strategies. *Computers & Education*, 59(2), 750–761. <https://doi.org/10.1016/j.compedu.2012.02.005>
- Shin, N., Jonassen, D. H., & McGee, S. (2003). Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching*, 40(1), 6–33. <https://doi.org/10.1002/tea.10058>
- Singer, F. M., & Voica, C. (2013). A problem-solving conceptual framework and its implications in designing problem-posing tasks. *Educational Studies in Mathematics*, 83(1), 9–26. <https://doi.org/10.1007/s10649-012-9422-x>
- Smith, T. J., Pasero, S. L., & McKenna, C. M. (2014). Gender Effects on Student Attitude Toward Science. *Bulletin of Science, Technology & Society*, 34(1–2), 7–12. <https://doi.org/10.1177/0270467614542806>
- Soeharto, S., & Csapó, B. (2022). Assessing Indonesian student inductive reasoning: Rasch analysis. *Thinking Skills and Creativity*, 46, 101132. <https://doi.org/10.1016/j.tsc.2022.101132>
- Song, H.-D., & Grabowski, B. L. (2006). Stimulating Intrinsic Motivation for Problem Solving Using Goal-Oriented Contexts and Peer Group Composition. *Educational Technology Research and Development*, 54(5), 445–466. <https://doi.org/10.1007/s11423-006-0128-6>
- Soto-Ardila, L. M., Caballero-Carrasco, A., & Casas-García, L. M. (2022). Teacher expectations and students' achievement in solving elementary arithmetic problems.

- Heliyon*, 8(5), e09447. <https://doi.org/10.1016/j.heliyon.2022.e09447>
- Stadler, M., Becker, N., Gödker, M., Leutner, D., & Greiff, S. (2015). Complex problem solving and intelligence: A meta-analysis. *Intelligence*, 53, 92–101. <https://doi.org/10.1016/j.intell.2015.09.005>
- Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 36(5), 404–411. <https://doi.org/https://doi.org/10.2307/30034943>
- Sternberg, R. (2018). Speculations on the Role of Successful Intelligence in Solving Contemporary World Problems †. *Journal of Intelligence*, 6(1), 4. <https://doi.org/10.3390/jintelligence6010004>
- Sternberg, R. J., & Kalmar, D. A. (2004). What do we know about the nature of reasoning. In *The nature of reasoning* (pp. 443–455). Cambridge University Press.
- Strayer, F. F., Verissimo, M., Vaughn, B. E., & Howes, C. (1995). A Quantitative Approach to the Description and Classification of Primary Social Relationships. *Monographs of the Society for Research in Child Development*, 60(2/3), 49. <https://doi.org/10.2307/1166170>
- Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, 63, 43–51. <https://doi.org/10.1016/j.compedu.2012.11.019>
- Sung, H., & Hwang, G. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, 63, 43–51. <https://doi.org/10.1016/j.compedu.2012.11.019>
- Susilawati, S., Aznam, N., & Paidi, P. (2022). Attitudes Towards Science: A Study of Gender Differences and Grade Level. *European Journal of Educational Research*, volume-11- (volume-11-issue-2-april-2022), 599–608. <https://doi.org/10.12973/eu-jer.11.2.599>
- Swanson, H. L. (2004). Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Memory & Cognition*, 32(4), 648–661. <https://doi.org/10.3758/BF03195856>
- Szeberényi, J. (2013). Problem-solving test: Conditional gene targeting using the Cre/loxP recombination system. *Biochemistry and Molecular Biology Education*, 41(6), 445–449. <https://doi.org/10.1002/bmb.20731>
- Szeberényi, J. (2014). Problem-solving test: E2F: A family of S phase regulating transcription factors. *Biochemistry and Molecular Biology Education*, 42(1), 88–90. <https://doi.org/10.1002/bmb.20753>
- Taghizadeh, M., & Hajhosseini, F. (2021). Investigating a Blended Learning Environment: Contribution of Attitude, Interaction, and Quality of Teaching to Satisfaction of Graduate Students of TEFL. *The Asia-Pacific Education Researcher*, 30(5), 459–469. <https://doi.org/10.1007/s40299-020-00531-z>
- Teo, T., & Noyes, J. (2011). An assessment of the influence of perceived enjoyment and attitude on the intention to use technology among pre-service teachers: A structural equation modeling approach. *Computers & Education*, 57(2), 1645–1653. <https://doi.org/10.1016/j.compedu.2011.03.002>
- Toma, R. B., & Greca, I. M. (2018). The Effect of Integrative STEM Instruction on

- Elementary Students' Attitudes toward Science. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(4). <https://doi.org/10.29333/ejmste/83676>
- Tomic, W. (1995). Training in Inductive Reasoning and Problem Solving. *Contemporary Educational Psychology*, 20(4), 483–490. <https://doi.org/10.1006/ceps.1995.1036>
- Tsai, C.-Y. (2015). Improving Students' PISA Scientific Competencies Through Online Argumentation. *International Journal of Science Education*, 37(2), 321–339. <https://doi.org/10.1080/09500693.2014.987712>
- Tuan \*, H., Chin, C., & Shieh, S. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education*, 27(6), 639–654. <https://doi.org/10.1080/0950069042000323737>
- Uluçınar Sağır, Ş. (2012). The Primary School Students' Attitude and Anxiety Towards SCience. *Journal of Baltic Science Education*, 11(2), 127–140. <https://doi.org/10.33225/jbse/12.11.127>
- van Aalderen-Smeets, S. I., Walma van der Molen, J. H., & Asma, L. J. F. (2012). Primary teachers' attitudes toward science: A new theoretical framework. *Science Education*, 96(1), 158–182. <https://doi.org/10.1002/sce.20467>
- van Aalderen-Smeets, S., & Walma van der Molen, J. (2013). Measuring Primary Teachers' Attitudes Toward Teaching Science: Development of the Dimensions of Attitude Toward Science (DAS) Instrument. *International Journal of Science Education*, 35(4), 577–600. <https://doi.org/10.1080/09500693.2012.755576>
- van de Vijver, F. J. R. (2002). Inductive reasoning in Zambia, Turkey, and the Netherlands. *Intelligence*, 30(4), 313–351. [https://doi.org/10.1016/S0160-2896\(02\)00084-3](https://doi.org/10.1016/S0160-2896(02)00084-3)
- Van Der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., Kramer, J., Warmuth, E., Heekeren, H. R., & Wartenburger, I. (2008). Resource allocation and fluid intelligence: Insights from pupillometry. *Psychophysiology*, 47(1), 158–169. <https://doi.org/https://doi.org/10.1111/j.1469-8986.2009.00884.x>
- van Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning. *Contemporary Educational Psychology*, 58, 260–275. <https://doi.org/10.1016/j.cedpsych.2019.03.005>
- Van Vo, D., & Csapó, B. (2020). Development of inductive reasoning in students across school grade levels. *Thinking Skills and Creativity*, 37, 100699. <https://doi.org/10.1016/j.tsc.2020.100699>
- Van Vo, D., & Csapó, B. (2021). Exploring students' science motivation across grade levels and the role of inductive reasoning in science motivation. *European Journal of Psychology of Education*. <https://doi.org/10.1007/s10212-021-00568-8>
- Van Vo, De, & Csapó, B. (2023). Exploring Inductive Reasoning, Scientific Reasoning and Science Motivation, and Their Role in Predicting STEM Achievement Across Grade Levels. *International Journal of Science and Mathematics Education*. <https://doi.org/10.1007/s10763-022-10349-4>
- Vandecandelaere, M., Speybroeck, S., Vanlaar, G., De Fraine, B., & Van Damme, J. (2012). Learning environment and students' mathematics attitude. *Studies in Educational Evaluation*, 38(3–4), 107–120. <https://doi.org/10.1016/j.stueduc.2012.09.001>

- Vedder-Weiss, D., & Fortus, D. (2012). Adolescents' declining motivation to learn science: A follow-up study. *Journal of Research in Science Teaching*, *49*(9), 1057–1095. <https://doi.org/10.1002/tea.21049>
- Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, *14*(1), 89–109. <https://doi.org/10.1016/j.learninstruc.2003.10.004>
- Vieira, R. M., & Tenreiro-Vieira, C. (2016). Fostering Scientific Literacy and Critical Thinking in Elementary Science Education. *International Journal of Science and Mathematics Education*, *14*(4), 659–680. <https://doi.org/10.1007/s10763-014-9605-2>
- Vincent, A. S., Decker, B. P., & Mumford, M. D. (2002). Divergent Thinking, Intelligence, and Expertise: A Test of Alternative Models. *Creativity Research Journal*, *14*(2), 163–178. [https://doi.org/10.1207/S15326934CRJ1402\\_4](https://doi.org/10.1207/S15326934CRJ1402_4)
- Vollmer, F., & Kaufmann, G. (1975). Achievement motivation and problem solving. *Scandinavian Journal of Psychology*, *16*(1), 323–326. <https://doi.org/10.1111/j.1467-9450.1975.tb00200.x>
- Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction*, *10*(4), 293–309. [https://doi.org/10.1016/S0959-4752\(99\)00031-6](https://doi.org/10.1016/S0959-4752(99)00031-6)
- Walker, F., Link, F., & Nickolaus, N. (2016). A multidimensional structure of domain-specific problem-solving competencies of electronics technicians for automaton technology. *Empirical Research in Vocational Education and Training*, *8*(7), 1–26. <https://doi.org/10.1186/s40461-016-0034-z>
- Wang, M., Wu, B., Kinshuk, Chen, N.-S., & Spector, J. M. (2013). Connecting problem-solving and knowledge-construction processes in a visualization-based learning environment. *Computers & Education*, *68*, 293–306. <https://doi.org/10.1016/j.compedu.2013.05.004>
- Wang, W., Qing, X., Wang, J., He, T., Fan, R., & Huang, Y. (2022). Bioaccumulation and potential risk of organophosphate flame retardants in coral reef fish from the Nansha Islands, South China Sea. *Chemosphere*, *287*, 132125. <https://doi.org/10.1016/j.chemosphere.2021.132125>
- Waschl, N., & Burns, N. R. (2020). Sex differences in inductive reasoning: A research synthesis using meta-analytic techniques. *Personality and Individual Differences*, *164*, 109959. <https://doi.org/10.1016/j.paid.2020.109959>
- Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2020). The moderating effect of prior knowledge on the relationship between intelligence and complex problem solving – Testing the Elshout-Raaheim hypothesis. *Intelligence*, *83*, 101502. <https://doi.org/10.1016/j.intell.2020.101502>
- Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences*, *35*(4), 863–875. [https://doi.org/10.1016/S0191-8869\(02\)00288-X](https://doi.org/10.1016/S0191-8869(02)00288-X)
- Wendt, J. L., & Rockinson-Szapkiw, A. (2018). A psychometric evaluation of the english version of the dimensions of attitudes toward science instrument with a U.S. population of elementary educators. *Teaching and Teacher Education*, *70*, 24–33.

<https://doi.org/10.1016/j.tate.2017.11.009>

- Wicaksono, A. G. C., & Korom, E. (2022a). The measurement of inductive reasoning skill in higher education in Indonesia. *JURE 2022: Unpredictable Challenges - Education in a Rapidly Changing World*.
- Wicaksono, A. G. C., & Korom, E. (2022b). Review of problem-solving measurement: an assessment developed in the Indonesian context. *Participatory Educational Research*, 9(1), 116–136. <https://doi.org/10.17275/per.22.7.9.1>
- Wicaksono, A. G. C., & Korom, E. (2023a). Attitudes towards science in higher education: Validation of questionnaire among science teacher candidates and engineering students in Indonesia. *Heliyon*, 9(9), e20023. <https://doi.org/10.1016/j.heliyon.2023.e20023>
- Wicaksono, A. G. C., & Korom, E. (2023b). Role of inductive reasoning, gender, learning satisfaction, and educational and career preference in predicting scientific competency in high school. *Thinking Skills and Creativity*, 49, 101376. <https://doi.org/10.1016/j.tsc.2023.101376>
- Wicaksono, A. G. C., & Korom, E. (2023c). The profile of high school students' problem-solving ability in Indonesia. *XIX 19th Conference on Educational Assessment : PÉK 2023*, 26. [https://www.edu.u-szeged.hu/pek2023/download/PEK\\_2023\\_CEA\\_2023\\_absztraktkotet.pdf#page=27](https://www.edu.u-szeged.hu/pek2023/download/PEK_2023_CEA_2023_absztraktkotet.pdf#page=27)
- Wicaksono, A. G. C., & Korom, E. (2022c). The relationship between science motivation, attitude toward science, and inductive reasoning among science major students in Indonesia. *XVIII 18th Conference on Educational Assessment*, 22.
- Wicaksono, A. G. C., & Korom, E. (2021). The validation of attitude towards science questionnaire among Indonesian undergraduate students. *EARLI 2021: Education and Citizenship: Learning and Instruction and the Shaping of Futures*, 27.
- Wirth, J., & Klieme, E. (2003). Computer-based Assessment of Problem Solving Competence. *Assessment in Education: Principles, Policy & Practice*, 10(3), 329–345. <https://doi.org/10.1080/0969594032000148172>
- Witkin, H. A., & Goodenough, D. R. (1981). Cognitive styles: essence and origins. Field dependence and field independence. *Psychological Issues*, 51, 1–141. <http://europemc.org/abstract/MED/7012884>
- Wittmann, W. W., & Hatrup, K. (2004). The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioral Science*, 21(4), 393–409. <https://doi.org/10.1002/sres.653>
- Wolff, K. (2017). Engineering problem-solving knowledge: the impact of context. *Journal of Education and Work*, 30(8), 840–853. <https://doi.org/10.1080/13639080.2017.1380299>
- Wood, E. M., Pidgeon, A. M., Liu, F., & Mladenoff, D. J. (2012). Birds see the trees inside the forest: The potential impacts of changes in forest composition on songbirds during spring migration. *Forest Ecology and Management*, 280, 176–186. <https://doi.org/10.1016/j.foreco.2012.05.041>
- Wu, H., & Molnár, G. (2018). Interactive problem solving: assessment and relations to combinatorial and inductive reasoning. *Journal of Psychological and Educational Research*, 26(1), 90–105. <http://www.marianjournals.com/book/issue-1-volume-26-2018/>

- Wu, H., & Molnár, G. (2021). Logfile analyses of successful and unsuccessful strategy use in complex problem-solving: a cross-national comparison study. *European Journal of Psychology of Education, 36*(4), 1009–1032. <https://doi.org/10.1007/s10212-020-00516-y>
- Wu, H., & Molnár, G. (2022). Analysing Complex Problem-Solving Strategies from a Cognitive Perspective: The Role of Thinking Skills. *Journal of Intelligence, 10*(3), 46. <https://doi.org/10.3390/jintelligence10030046>
- Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving — More than reasoning? *Intelligence, 40*(1), 1–14. <https://doi.org/10.1016/j.intell.2011.11.003>
- Wüstenberg, S., Greiff, S., Molnár, G., & Funke, J. (2014). Cross-national gender differences in complex problem solving and their determinants. *Learning and Individual Differences, 29*, 18–29. <https://doi.org/10.1016/j.lindif.2013.10.006>
- Yang, Y.-T. C., & Wu, W.-C. I. (2012). Digital storytelling for enhancing student academic achievement, critical thinking, and learning motivation: A year-long experimental study. *Computers & Education, 59*(2), 339–352. <https://doi.org/10.1016/j.compedu.2011.12.012>
- Yeh, F.-Y., Tran, N.-H., Hung, S. H., & Huang, C.-F. (2022). A Study of Environmental Literacy, Scientific Performance, and Environmental Problem-Solving. *International Journal of Science and Mathematics Education, 20*(8), 1883–1905. <https://doi.org/10.1007/s10763-021-10223-9>
- Yeh, Y., Lin, T., Hsu, Y., Wu, H., & Hwang, F. (2015). Science Teachers' Proficiency Levels and Patterns of TPACK in a Practical Context. *Journal of Science Education and Technology, 24*(1), 78–90. <https://doi.org/10.1007/s10956-014-9523-7>
- You, H. S., Kim, K., Black, K., & Min, K. W. (2018). Assessing Science Motivation for College Students: Validation of the Science Motivation Questionnaire II using the Rasch-Andrich Rating Scale Model. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(4). <https://doi.org/10.29333/ejmste/81821>
- Yu, W., She, H., & Lee, Y. (2010). The effects of Web-based/non-Web-based problem-solving instruction and high/low achievement on students' problem-solving ability and biology achievement. *Innovations in Education and Teaching International, 47*(2), 187–199. <https://doi.org/10.1080/14703291003718927>
- Zakaria, E., & Ngah, N. (2011). A Preliminary Analysis of Students' Problem-Posing Ability and its Relationship to Attitudes Towards Problem Solving. *Research Journal of Applied Sciences, Engineering and Technology, 3*(9), 866–870. <https://maxwellsci.com/print/rjaset/v3-866-870.pdf>
- Zandalinas, S. I., Fritschi, F. B., & Mittler, R. (2021). Global Warming, Climate Change, and Environmental Pollution: Recipe for a Multifactorial Stress Combination Disaster. *Trends in Plant Science, 26*(6), 588–599. <https://doi.org/10.1016/j.tplants.2021.02.011>
- Zhang, D., & Campbell, T. (2011). The Psychometric Evaluation of a Three-Dimension Elementary Science Attitude Survey. *Journal of Science Teacher Education, 22*(7), 595–612. <https://doi.org/10.1007/s10972-010-9202-3>
- Zhao, N., Teng, X., Li, W., Li, Y., Wang, S., Wen, H., & Yi, M. (2019). A path model for metacognition and its relation to problem-solving strategies and achievement for

different tasks. *ZDM*, *51*(4), 641–653. <https://doi.org/10.1007/s11858-019-01067-3>

Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of Experimental Child Psychology*, *110*(4), 481–498. <https://doi.org/10.1016/j.jecp.2011.06.001>

Zhou, Q., Hirst, G., & Shipton, H. (2012). Promoting Creativity at Work: The Role of Problem-Solving Demand. *Applied Psychology*, *61*(1), 56–80. <https://doi.org/10.1111/j.1464-0597.2011.00455.x>

## APPENDIX A. MicroDYN test

### Knowledge acquisition phase

#### 1. Cat

Find out about the relations and plot them in the model. Please note: The clock will be ticking while you are looking at this page. 178

Brekon

Mikas

Purring

Movement

Help

Reset

Apply

Brekon

Purring

Mikas

Movement

Next

#### 2. Drawing

Find out about the relations and plot them in the model. Please note: The clock will be ticking while you are looking at this page. 178

Topax

Floba

Brightness

Fluidity

Help

Reset

Apply

Topax

Brightness

Floba

Fluidity

Next

#### 3. Petrol

Find out about the relations and plot them in the model.

175

Carenol

Noresal

Farunin

Speed

Exhaust emissions

Help

Reset

Apply

Carenol

Noresal

Farunin

Speed

Exhaust emissions

Next

#### 4. Games

Find out about the relations and plot them in the model.

177

Blue chips

Green chips

Red chips

Cards

Pawns

Score

Reset

Help

Apply

Blue chips

Green chips

Red chips

Cards

Pawns

Score

Next

## 5. Fruit scent

Find out about the relations and plot them in the model.

178

Norilan

Miral

Carumin

Fresh

Fruity

Flowery

Help

Reset

Apply

Norilan

Miral

Carumin

Fresh

Fruity

Flowery

Next

## Knowledge application phase

### 1. Cats

Reach the given target area in no more than four steps. Please note: The clock will be ticking while you are looking at this page.

86

Brekon

Mikas

Purring

Movement

Help

Reset

Apply

Brekon

Mikas

Purring

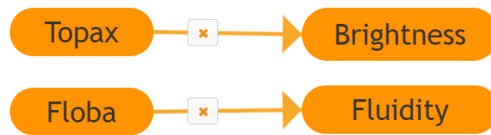
Movement

Next

## 2. Drawing

Reach the given target area in no more than four steps. Please note: The clock will be ticking while you are looking at this page.

87

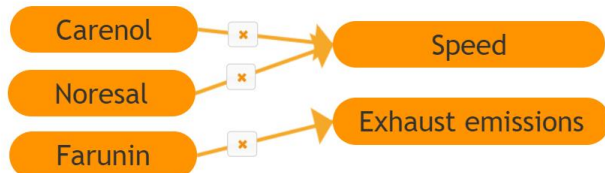
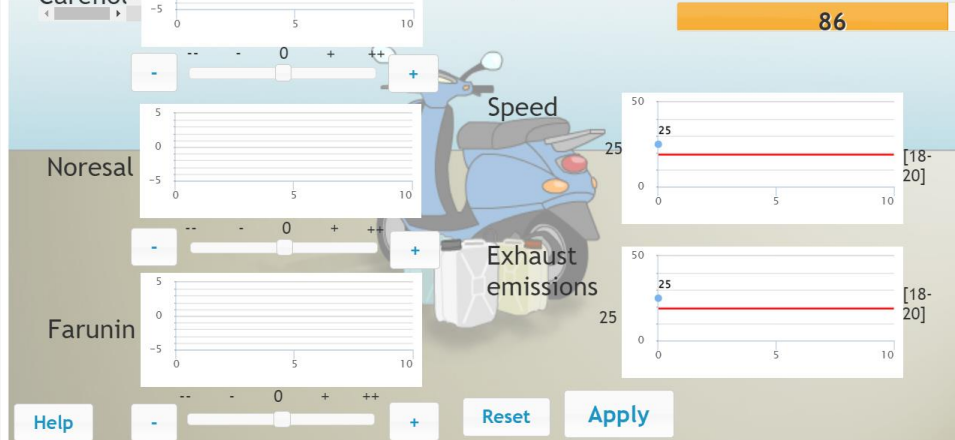


Next

## 3. Petrol

Reach the given target area in no more than four steps.

86



Next

## 4. Games

Reach the given target area in no more than four steps.

88

Blue chips

Green chips

Red chips

Cards

Pawns

Score

Reset

Apply

Help

Blue chips

Green chips

Red chips

Cards

Pawns

Score

Next

## 5. Fruit scent

Reach the given target area in no more than four steps.

86

Norilan

Miral

Carumin

Fresh

Fruity

Flowery

Reset

Apply

Help

Norilan

Miral

Carumin

Fresh

Fruity

Flowery

Next

## APPENDIX B. Science Problem-solving test

### Science problem-solving test

This test consists of 11 multiple choice questions and measuring the problem-solving skills in natural science topic. Each question is provided by five answers/choices.

Choose the correct answer!

### *Ecosystem and environmental issue*

The scientists identified the level of heavy metal pollution) in five different rivers in Sumatra Island, Indonesia. To classify the level of pollution, they use a physical indicator (temperature), a chemical indicator (pH and sulphate), and a bioindicator. In terms of bioindicators, they identify the present zooplankton and indicate the high level of metal pollution based on the frequency of several zooplankton genera. The result of the pollution analysis is shown in Table 1.

Table 1. The result of water pollution identity in the Sumatra River

	Indicator	River				
		A	B	C	D	E
Physical	Temperature (°C)	25	27	26	26	26
	pH	6.06	7.43	6.85	6.25	7.01
Chemical	Pb (mg L <sup>-1</sup> )	0.008	0.024	0.014	0.006	0.017
	Mn (mg L <sup>-1</sup> )	0.081	0.42	0.228	0.023	0.226
	Zn (mg L <sup>-1</sup> )	0.039	0.071	0.055	0.014	0.075
	<i>Anuraeopsis sp</i>	+	+++	++	+	+++
	<i>Brachionus caudatus</i>	++	++	++	++	+
Biology	<i>Trichocerca gracilis</i>	+	+	-	+	+
	<i>Keratella cochlearis</i>	+	+++	+++	-	+++
	<i>Arcella megastoma</i>	++	-	+	-	-

Note: the frequency of zooplankton is noted as very frequent (+++), frequent (++), infrequent (+), and does not exist (-).

1. What are the best factors to indicate the high level of water pollution in the river?
  - a. high pH value, high chemical concentration (Pb, Mn, and Zn), and high number of *Brachionus*
  - b. low pH value, high chemical concentration (Pb, Mn, and Zn), and high number of *Anuraeopsis*
  - c. neutral pH value, high chemical concentration (Pb, Mn, and Zn), and high number of *Trichocerca*
  - d. high pH value, high chemical concentration (Pb, Mn, and Zn), and high number of *Keratinella*
  - e. low pH value, high chemical concentration (Pb, Mn, and Zn), and high number of *Arcella*

### *Bioremediation*

The heavy metal pollution in river like lead (Pb), manganese (Mn) and Zinc (Zn), threatens both the ecosystem and human health because of the high toxicity. The scientists try to use bioremediation for heavy metal with microbial agent (*Mucor circinelloides*). They do the laboratory research and found that the activity of bioremediation is affected by several factor including initial inoculation amount, temperature, and pH (Figure 1).

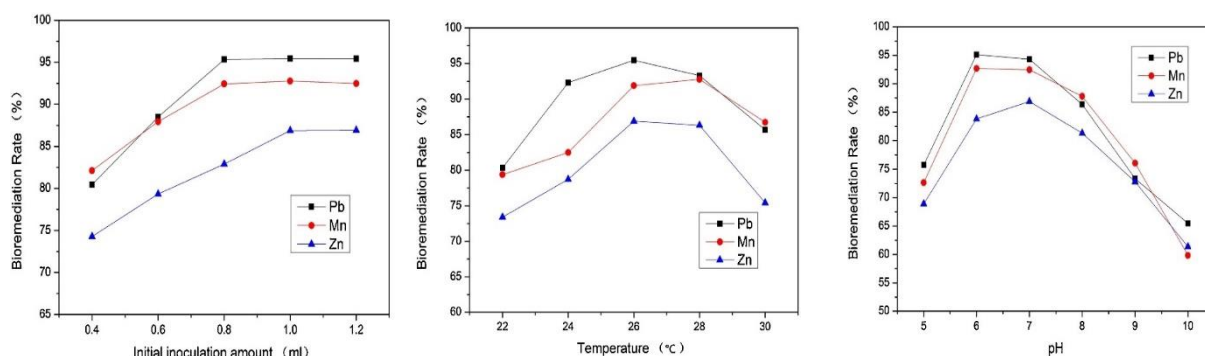


Figure 1. The bioremediation rate in different parameters

2. If the researcher introduced the microbial agent culture in the river E ( $T= 26^{\circ}\text{C}$ ,  $\text{pH}= 7.01$ ). How is the bioremediation rate that can be predicted for water pollution?

- The bioremediation rate for Zn will reach 95%.
- The bioremediation rate for Zn will reach 85%.
- The bioremediation rate for Mn will reach 85%.
- The bioremediation rate for Mn will reach 80%.
- The bioremediation rate for Pb will reach 75%.

### Migration

The migration of songbirds (*Oreothlypis peregrina*) is affected by the formation of forest and trees. The songbirds primarily use trees as foraging substrates during spring migration. They prey on invertebrates living on and around the tree. Scientists observed several forests in the US and revealed some aspects of the tree that influence the songbird behavior and migration preference (Table 2).

Table 2. Tree species preference for songbird migration based on foraging and leaf petiole length.

Tree species	Scientific name	Shading type	Foraging index (M)	Leaf petiole length (cm)	Migration-tree preference
A	<i>Populus grandidentata</i>	Shade intolerant	0.04	3.01	1.61
B	<i>Betula papyrifera</i>	Shade intolerant	0.14	1.19	-1.09
C	<i>Carya cordiformis</i>	Intermediate shade tolerant	0.04	3.78	-7.58
D	<i>Fraxinus americana</i>	Intermediate shade tolerant	0.26	0.35	0.85
E	<i>Quercus rubra</i>	Intermediate shade tolerant	0.13	1.75	26.84
F	<i>Quercus alba</i>	Intermediate shade tolerant	0.16	0.82	8.8
G	<i>Ulmus americana</i>	Intermediate shade tolerant	0.15	0.2	4.76
H	<i>Ulmus rubra</i>	Intermediate shade tolerant	0.29	0.22	4.18
I	<i>Acer rubrum</i>	Shade tolerant	0.03	4.8	-5.62
J	<i>Tilia americana</i>	Shade tolerant	0.02	5	-8.08
K	<i>Acer saccharum</i>	Shade tolerant	0.02	4.75	-15.87

Note: in migration-tree preference aspect, positive value show preference and negative value show aversion. Left petiole length above 1 cm is long and below 1 cm is short (adapted from Wood et al., 2012)

3. What is the songbird's preference of the tree during migration?
  - a. The songbirds prefer shade tolerant tree during migration. They are mostly foraging substrates on the tree with long petiole leaf.
  - b. The songbirds prefer shade tolerant tree during migration. They are mostly foraging substrates on the tree with short petiole leaf.
  - c. The songbirds prefer shade intolerant tree during migration. They are mostly foraging substrates on the tree with short petiole leaf.
  - d. The songbirds prefer intermediate shade tolerant tree during migration. They are mostly foraging substrates on the tree with long petiole leaf.
  - e. The songbirds prefer intermediate shade tolerant tree during migration. They are mostly foraging substrates on the tree with short petiole leaf.
  
4. If the Government and scientists want to do deforestation to helps the songbird migration, which tree species that suitable for the forest composition?
  - a. B, I, F, C, and K
  - b. J, K, A, B, and G
  - c. D, J, E, F, and I
  - d. D, E, F, G, and H
  - e. D, G, H, K, and C

**Pathogen – Fungi**

The Department of Plant Pathology identifies the most common fungal disease that attacks tomato plants and summarized the symptom in Table 3.

Table 3. The identification result of fungal disease in plant (adapted from Mark & Brooke (2006))

Name of Fungal	Optimum habitat		Symptom
	T (°C)	Humidity (%)	
<i>Septoria lycopersici</i>	17-31	50-60	Circular spots on leaf about 1/8 inch in diameter. The lesions gradually develop grayish white centers with dark edges.
<i>Alternaria solani</i>	20-34	45-70	Premature loss of lower leaves
<i>Colletotrichum coccodes</i>	19-36	30-55	Visible on ripe or ripening fruit as small, circular, indented spots in the skin. As these spots expand, they develop dark centers or concentric rings of dark specks.
<i>Fusarium oxysporum</i>	25-35	40-60	Leaf yellowing and wilting that progress upward from the base of the stem
<i>Botrytis cinerea</i>	24-27	80-81	Light-gray fuzzy growth that appears on stems and leaves. Soft rot of the stem and the fruit

5. The farmer plant tomato in the field with the normal temperature (26°C) and low humidity (60-80%). 40 days later, he found the dark spot on several leaves in the upper part of plants. But the stem is in normal condition and the flower is developing. From which fungal infection that possibly cause the disease?

- a. *Septoria lycopersici*
- b. *Botrytis cinerea*
- c. *Fulvia fulva*
- d. *Colletotrichum coccodes*
- e. *Alternaria solani*

***Plant crossbreeding***

The scientists proved that the hybrid production of the new plant’s varieties of tomato can improve their resistance to fungal disease (i.e., powdery mildew) and improve the quality of the fruit production. There are several parental tomato plants that have potential to be involved in inbreeding process. In this case big fruit, long stem, and intolerant traits are dominant over small fruit, short stem, and resistance trait (Table 4).

Table 4. The traits of different tomato plants

Parental	Attributes (Traits)
TM1	Big fruit, long stem, intolerant to powdery mildew
TM2	Small fruit, long stem, intolerant to powdery mildew
TM3	Big fruit, long stem, resistance to powdery mildew
TM4	Small fruit, long stem, resistance to powdery mildew
TM5	Big fruit, short stem, intolerant to powdery mildew
TM6	Small fruit, short stem, intolerant to powdery mildew
TM7	Big fruit, short stem, resistance to powdery mildew
TM8	Small fruit, short stem, resistance to powdery mildew

6. Which parental plant that suitable for producing tomato that has big fruit, short stem, and high resistance of fungal disease?

- a. TM1 X TM3
- b. TM2 X TM4
- c. TM5 X TM6
- d. TM3 X TM7
- e. TM7 X TM8

***Exercise and energy***

Table 5 shows how many calories from food can be "burned" during exercise ().

Exercises for calories "burning"	Speed of exercise	Kcal "burned"/hour
Jogging	9 min/miles	700
Swimming	30 min/miles	540
Walking	20 min/miles	160

7. The adult male needs approximately 2500 kcal for their daily activity. When adult male eats food with 3200 kcal for a day. What kind of exercise they need to do to keep the calories balance for a day?

- a. Jogging for 30 minutes then walking for 1.5 hours
- b. Swimming for 1 hour then walking for 1 hour
- c. Swimming for 1 hour then jogging for 30 minutes
- d. Swimming for 1.5 hours
- e. Jogging for 2 hours

### Sea pollution

Coral reefs are foundational to marine ecosystem functioning and ecosystem service provision (Hoegh-Guldberg, 2011). It became the habitat of many organisms such as phytoplankton, coralline algae, sponges, gastropods, mollusk, corals, sea urchins, crabs, fish, and seals. Yet, in recent decades, coral reefs have been rapidly degrading in response to numerous anthropogenic stressors, including centuries of overfishing, global climate change, and environmental pollution. Among various pollutants, microplastics, as an emerging contaminant, have caused significant concern recently, because they are omnipresent within the marine environment, being found even in the most pristine areas of the planet, including the deep sea and polar region.

8. What the possible cause of the microplastic pollution abundance in the coral reef ecosystem?
- the gas emission from vehicle and factory near the beach
  - liquid waste of factory in coastal area
  - rubbish and waste from recreational activities at beach/sea
  - oil leak from industrial mining in the ocean
  - organic waste from industrial fishing in coastal area

### Microplastic

The microplastic residual in coral reef were identified by some scientist with measuring microplastic component such as tributyl phosphate (TBP), 2-ethylhexyl diphenyl phosphate (EDP), triphenyl phosphate (TPP) and tricresyl phosphate (TCP) from fish living in the coral reef. All those pollutants are toxic and cause damage to the endocrine system and reproductive ability of animals. The result of the microplastic pollutant in different fish species is given in Figure 2.

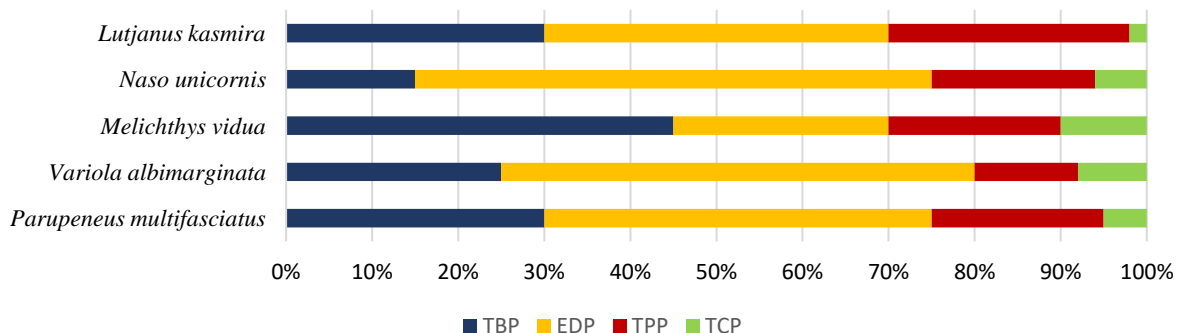


Figure 2. The level of microplastic residue in coral fish (adapted from Wang et al. (2022))

9. What is the most abundant microplastic pollutant found in all coral reefs? (put an order from the highest to the least)!
- TCP > TPP > TBP > EDP
  - EDP > TPP > TCP > TBP
  - TPP > TCP > EDP > TBP
  - EDP > TBP > TPP > TCP
  - TBP > EDP > TPP > TCP

### Conservation I

The endangered giant salamander (*Andrias davidianus*) is known to be historically distributed in rocky mountain streams and lakes with clear, running water, at moderate altitudes (below 1500 m, especially between 300 and 800 m), where the animals occupy hollows and cavities under water. Its wild populations and habitat have been declining due to habitat destruction and hunting for use in medicinal herbs and foods, and the species is currently critically endangered.

10. If the local government wants to make a conservation program for the giant salamander, which conservation is suitable for it?
- making their natural habitat as a conservation site
  - moving salamander in the different location far from human's residence
  - conserved the salamander into the zoo
  - moving all salamander into the breeding center together with other amphibians
  - making educational park close the salamander habitat, so they can live together with people

### Conservation II

This map below shows the area of Paju district, Korea. The blue polygon indicates the Hangang-Imjingang River. The red line indicates the nationally protected wetland area. The gray area indicates the region of interest for systematic conservation planning (Figure 3).

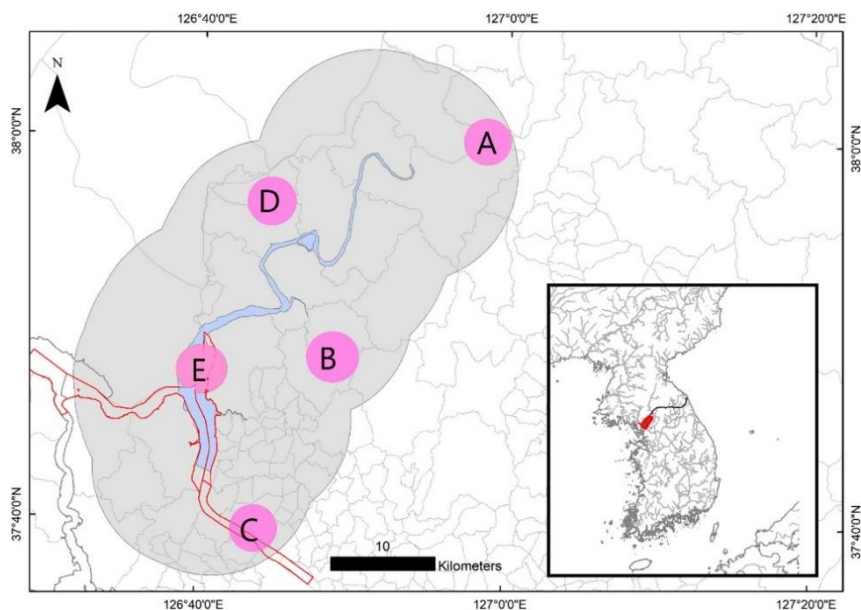


Figure 3. The conservation site of giant salamander (J. H. Kim et al., 2021)

11. Look at the map above! Among the pointed site (purple circle), which part is the best area for giant salamander conversation site?
- site A
  - site B
  - site C
  - site D
  - site E

### References

- Kim, J. H., Park, S., Kim, S. H., & Lee, E. J. (2021). Identifying high-priority conservation areas for endangered waterbirds using a flagship species in the Korean DMZ. *Ecological Engineering*, 159, 106080. <https://doi.org/10.1016/j.ecoleng.2020.106080>
- Mark, G. L., & Brooke, A. E. (2006). Tomato disease and disorder. In *Disease in outdoor production* (pp. 1–12). ISU Extension Publication.
- Wang, W., Qing, X., Wang, J., He, T., Fan, R., & Huang, Y. (2022). Bioaccumulation and potential risk of organophosphate flame retardants in coral reef fish from the Nansha Islands, South China Sea. *Chemosphere*, 287, 132125. <https://doi.org/10.1016/j.chemosphere.2021.132125>

Wood, E. M., Pidgeon, A. M., Liu, F., & Mladenoff, D. J. (2012). Birds see the trees inside the forest: The potential impacts of changes in forest composition on songbirds during spring migration. *Forest Ecology and Management*, 280, 176–186. <https://doi.org/10.1016/j.foreco.2012.05.041>

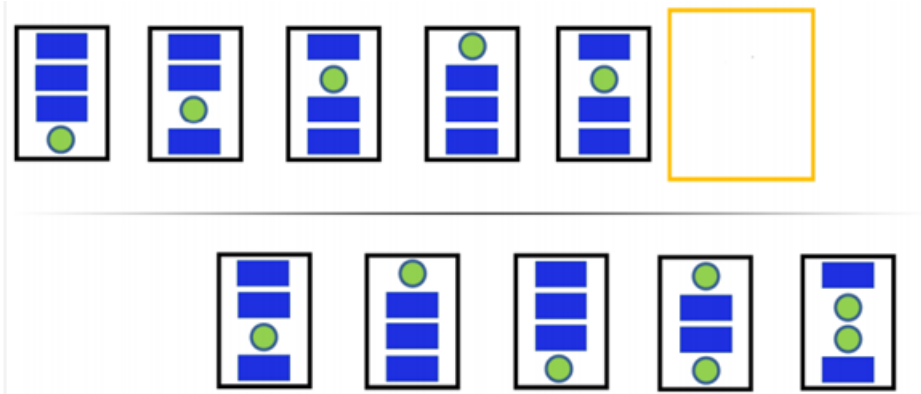
## APPENDIX C. Inductive Reasoning test

### Inductive reasoning test

The test contains 32 items and students has to complete the task by dragging the correct option (picture) into the box or type the correct number in the next empty box

#### Figure Series (FS)

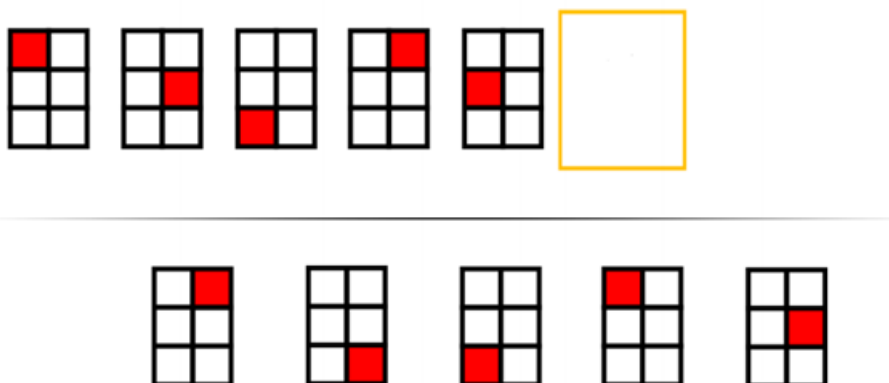
FS 1



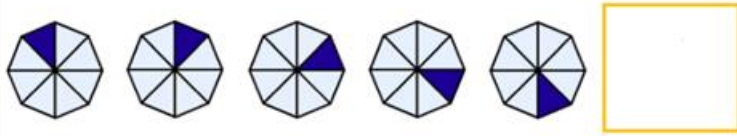
FS 2



FS 3



FS 4



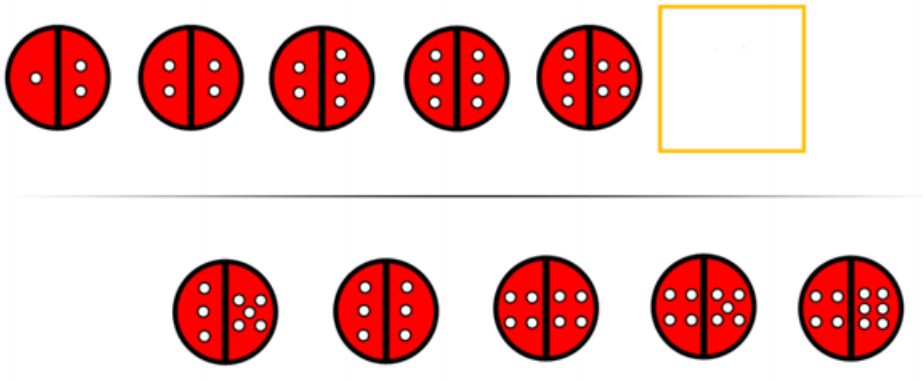
FS 5



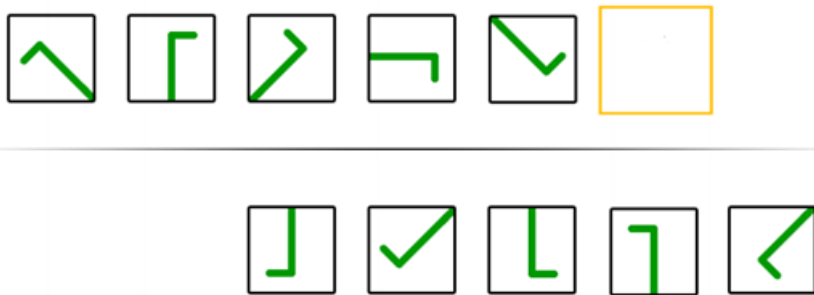
FS 6



FS 7

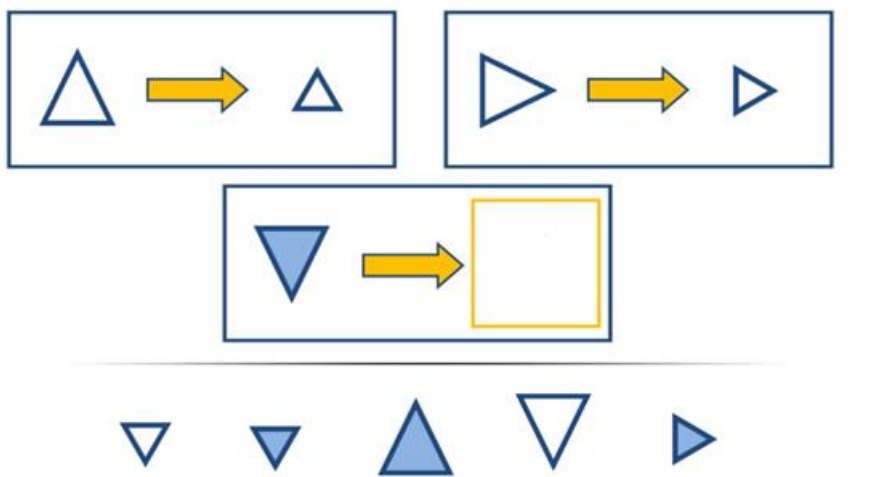


FS 8

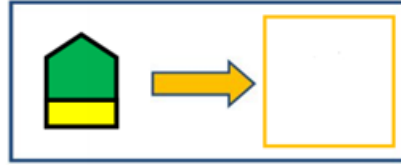
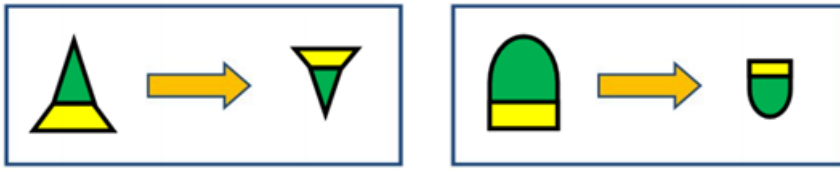


*Figure Analogy (FA)*

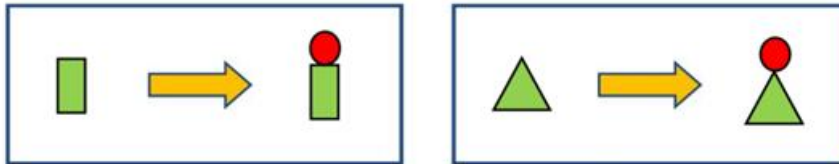
FA1



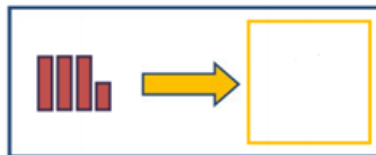
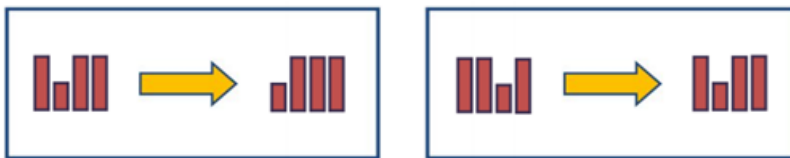
FA2



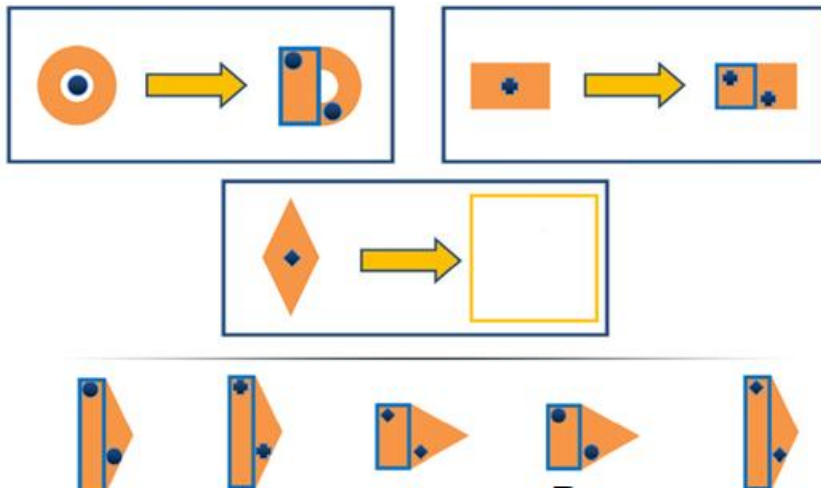
FA3



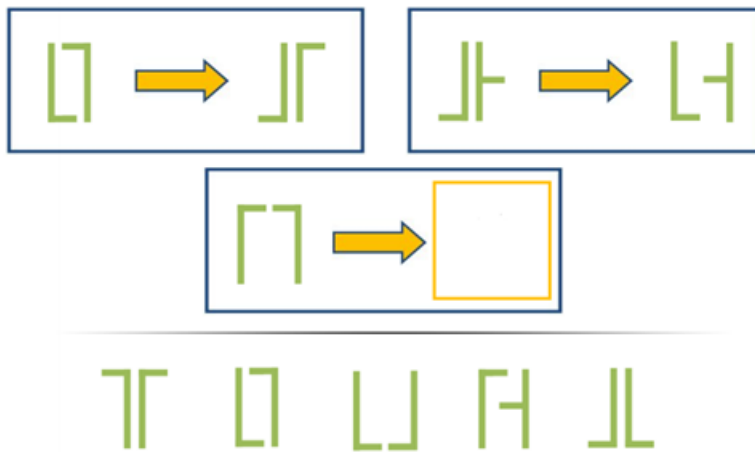
FA4



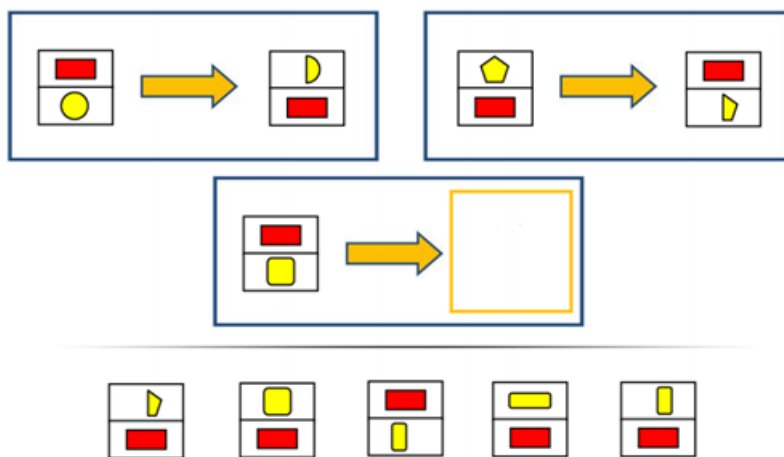
FA5



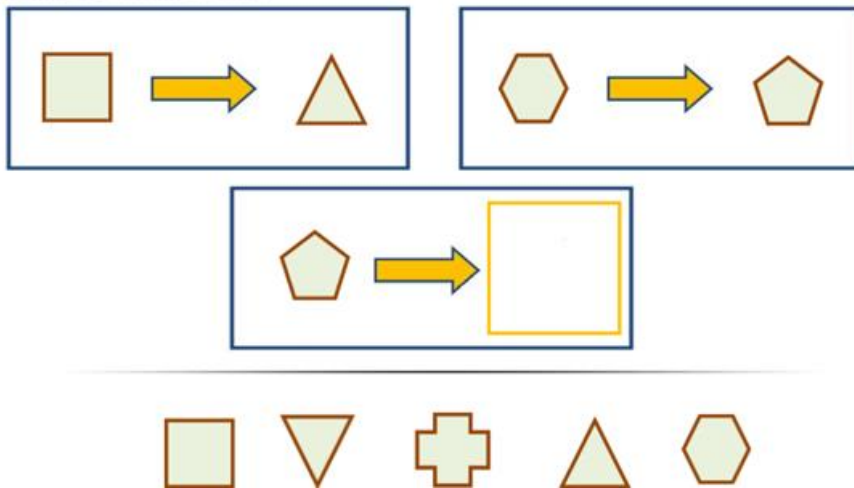
FA6



FA7

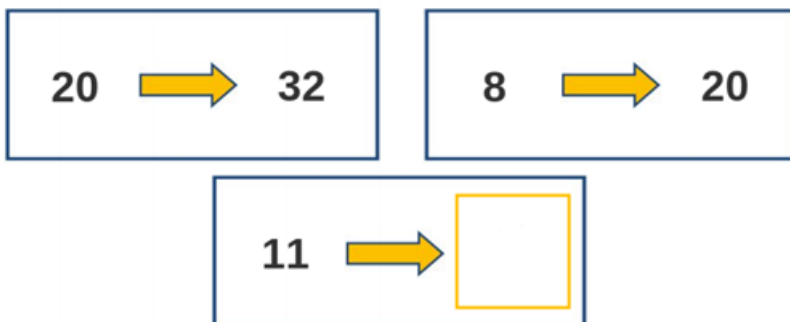


FA8

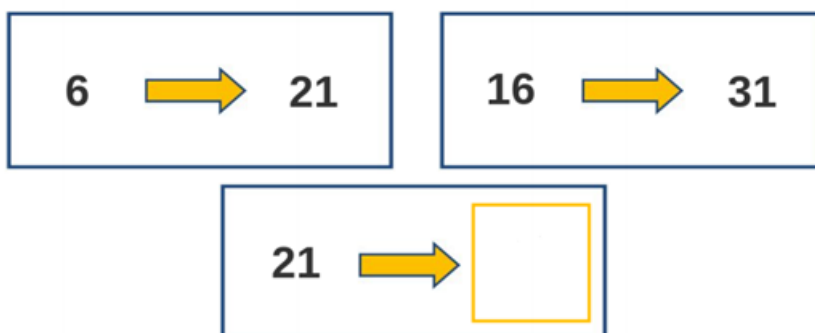


Number Analogy (NA)

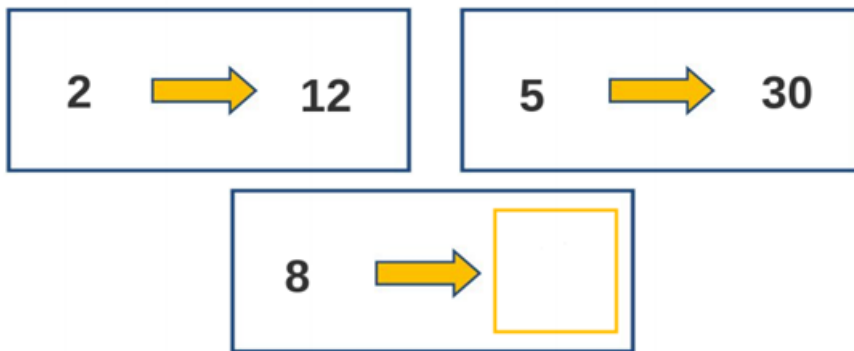
NA1



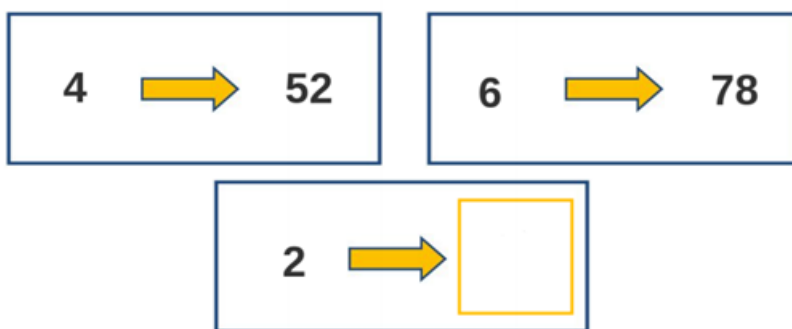
NA2



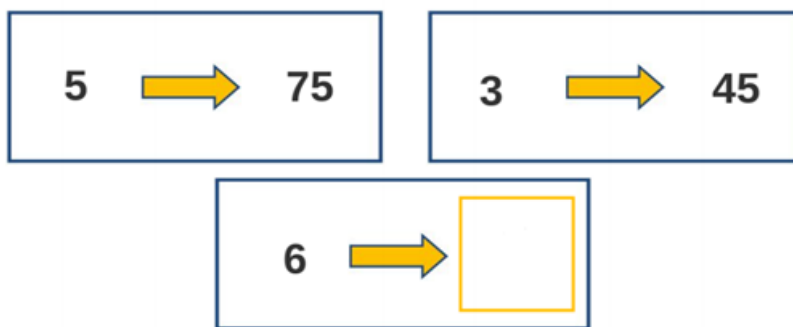
NA3



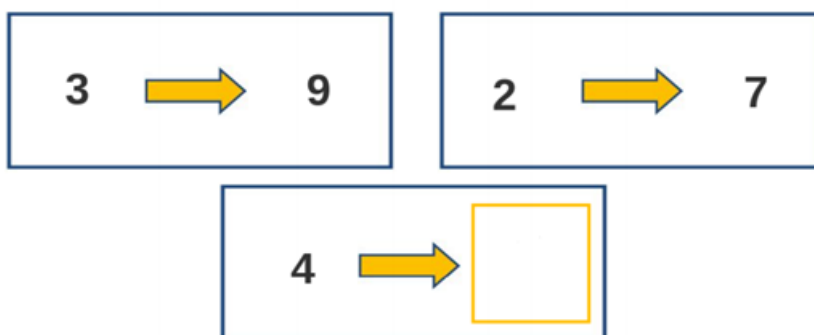
NA4



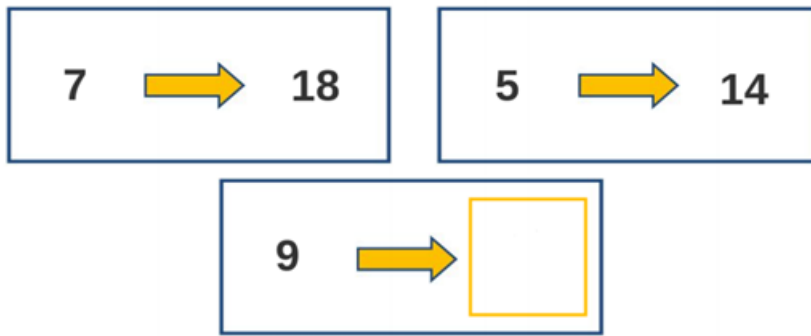
NA5



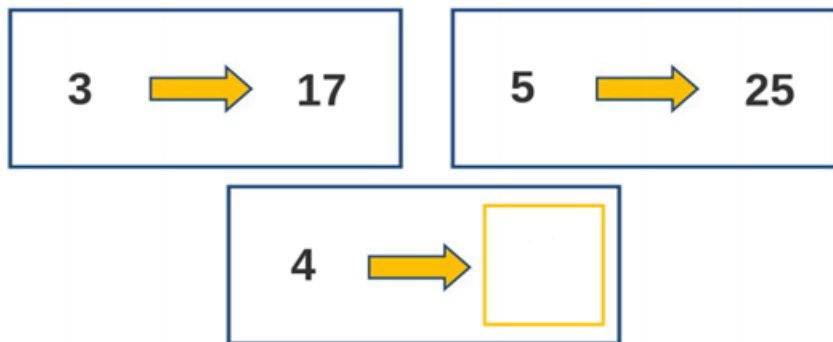
NA6



NA7



NA8



*Number Series (NS)*

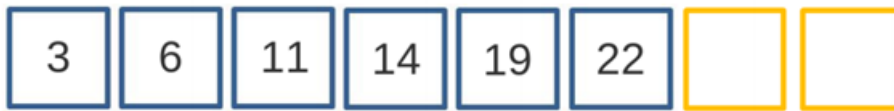
NS1



NS2



NS3



NS4



NS5



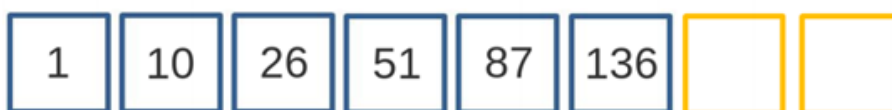
NS6



NS7



NS8



## APPENDIX D. Scientific Competency test

This test consists of 15 multiple choice questions and measures the reasoning ability in biology/natural science topic. Each question is provided with five answers/choices.

Choose the correct answer!

### *Mycorrhiza and legumes*

The association of mycorrhiza fungus with a host plant such as legumes is described as a long-lasting mutual adaptation. The development of mycorrhiza in the roots has some benefit for plants. They help to take immobile nutrients such as phosphate and zinc and acquire nitrate ( $\text{NO}_3^-$ ), ammonium ( $\text{NH}_4^+$ ), and amino acids (AAs) from the external medium/soil. However, symbiotic legumes with mycorrhiza give greater benefit to the plant compared to nonmycorrhiza legumes. Some laboratory research explored the differences between several legumes and the effect of their symbiotic relationship with mycorrhiza. They compared the effect of mycorrhiza, nonmycorrhiza and phosphate-rich soil on growth and nodule formation in legumes. The result is presented in Table 1.

Table 1. The effect of mycorrhiza and phosphate on the growth and nodulation of four legumes

Legumes	Total weight (g)			Nodulation		
	M	NM	P	M	NM	P
<i>Centrosema pubescens</i>	3.9	1.7	4.9	5	1	5
<i>Stylosanthes guyanensis</i>	1.6	0.5	0.9	5	0	5
<i>Trifolium repens</i>	2.6	1.6	4	5	1	5
<i>Lotus pedunculatus</i>	2	2.5	3.9	4	2	5

Note: M = with mycorrhiza, NM= without mycorrhiza, P= with additional 0.4 g phosphate molecules ( $\text{Ca}(\text{H}_2\text{PO}_4)_2 \cdot 2\text{H}_2\text{O}/\text{kg}$  soil), nodulation category from 0 = no nodule to 5 = abundance nodule.

1. What best describes the benefit of mycorrhiza symbiotic to legumes?
  - a. Absorption of the nutrient from the soil into the plants
  - b. improving plant photosynthesis rate and producing sugar molecules
  - c. synthesising the essential enzyme for plant growth
  - d. producing organic molecules that help plant against pathogens
  - e. increasing the water absorption from the soil
2. What happened to legumes that they cannot form a symbiotic with mycorrhiza and lack phosphate uptake from the soil?
  - a. Incomplete chlorophyll formation results in yellow leaves
  - b. decay of the root and lower stem
  - c. incomplete flower formation and small fruit size
  - d. abnormal growth in the stem and leaves
  - e. slow development and growth process (dwarf plants)

3. What is the effect of mycorrhiza and phosphate on nodule formation in legumes?
  - a. Nonmycorrhiza plants can produce nodules as much as mycorrhiza plants can.
  - b. the addition of phosphate does not improve nodule formation in legume plants
  - c. legume plants symbiotic with mycorrhiza have less ability to form nodules.
  - d. Treatment with mycorrhiza and phosphate improves nodule formation in legumes
  - e. mycorrhiza symbiotic will result in higher nodulation better than phosphate treatment.

### ***The Bird Migration***

Wind is an important factor to affect migrating birds' optimal flying speed and altitude. The speed and direction of the wind also correlate with the energy consumption during flight. During different times, the wind speed in a specific place is changing and combined with the topography of the migration route affects the seasonal migration pattern of birds. The wind speed record and the frequency of bird migration from August to October are presented in Figure 2.

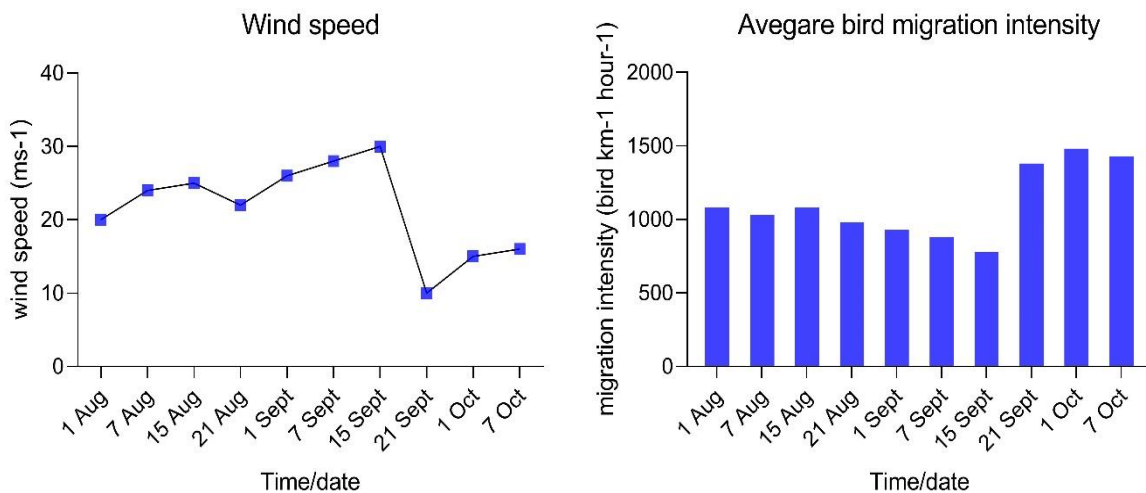


Figure 2. Wind speed and bird migration intensities

4. What is the relationship between wind speed and bird migration rate?
  - a. The change in wind speed does not affect the bird's migration rate.
  - b. The birds migrate in the period of low wind speed.
  - c. higher number of birds migrate during high wind speed;
  - d. the birds prefer high wind speed conditions for migration
  - e. there are few birds migrate during low wind speed
  
5. In what period does bird migration frequently occur?
  - a. early August
  - b. mid-August
  - c. early September
  - d. mid-September
  - e. early October

### **Biogas and Energy**

Biogas is a renewable source of energy produced by the decomposition of organic materials (feedstocks) under an oxygen-free and controlled temperature in the process known as anaerobic digestion with the help of microorganisms. The main components of biogas are methane (CH<sub>4</sub>), which makes it applicable as a natural gas. When the biogas units are combined with heat and power, it can produce various heating purposes and electricity. Instead, the constant temperature in the biogas process is important, a research on the ideal temperature for biogas was carried out and the results are shown in Figure 3.

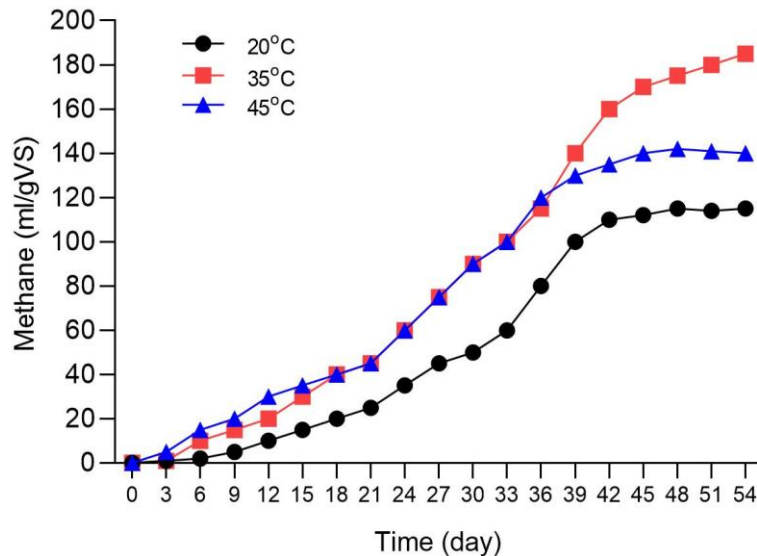


Figure 3. The effect of temperature for biogas methane production

6. Based on the figure above, which statements are correct for the effect of temperature on biogas methane production?

- Moderate temperature ( $T = 35^{\circ}\text{C}$ ) produces a lower methane gas compared to highest temperature ( $T = 45^{\circ}\text{C}$ )
- there is no difference in the effect of temperature for methane production during the first 30 days of biogas processing.
- The lower temperature ( $T = 20^{\circ}\text{C}$ ) has the lowest impact on the production of methane in biogas processing.
- Lower temperature ( $T = 20^{\circ}\text{C}$ ) produces methane gas with the same amount as moderate temperature ( $T = 35^{\circ}\text{C}$ )
- the highest temperature ( $T = 45^{\circ}\text{C}$ ) is more suitable for methane production in biogas processing;

7. What is the benefit of biogas for the sustainable environment?

- biogas uses microorganisms to produce gas and heat
- Biogas can be used for electricity production with less fuel consumption
- It does not require a lot of money to build a biogas plant.
- Biogas produces a large amount of methane gas
- biogas is applicable in small- and large-scale production.

### **Bioremediation**

Heavy metal environmental pollution has devastating effects on humans, plants, and animals. Chromium (Cr) is known for its widespread utility in various industries such as tanneries, chrome plating, chemical industry, mining, steel. Chromium pollution becomes a serious threat to the soil. It becomes harmful to soil microorganisms, inhibits nutrient uptake, and makes soil less fertile. The bioremediation strategy is used to minimise the level of chromium in the soil by bioagents such as *Streptomyces* sp. Research about Cr bioremediation is done by some researchers applying *Streptomyces* sp in polluted soil. Bioremediation activity is measured with the help of bioindicator plants such as *Zea mays* (corn). The level of Cr is calculated after 14 and 28 days in several parts of the plants (Table 2).

Table 2. Cr level in *Zea mays* after *Streptomyces* treatment

Cr accumulation in <i>Zea mays</i>	<i>Streptomyces griseus</i>			<i>Streptomyces lividans</i>		
	t-0	t-14	t-28	t-0	t-14	t-28
Root	6.2	3.4	0.3	5.6	1.8	1.8
Stem	10.3	1.8	0.4	9.8	4.2	1
leaf	9.9	1.9	1.1	7.8	3.6	0.8

Note: Metal accumulation in plants is in mg g<sup>-1</sup>; t= time (days)

8. Based on the data, which statements are correct about the result of bioremediation treatment? The lowest Cr accumulation is found in the...

- a. roots after 2 weeks of *Streptomyces* treatment
- b. stem after 2 weeks of *Streptomyces* treatment.
- c. leaf after 4 weeks of *Streptomyces* treatment
- d. root after 4 weeks of *Streptomyces* treatment
- e. stem after 4 weeks of *Streptomyces* treatment

9. Explain how fungal strains help in the chromium remediation process?

- a. absorbing the pollutant and accumulate it in their body.
- b. mobilising the chemical compound into their host plants;
- c. change the pH of their surrounding with the pollutant compound
- d. degrading the pollutant compound to a less toxic form
- e. produce mucus to trap the pollutant.

### **Global warming**

Global warming is a long-term heating system in the earth that is caused by the high amount of CO<sub>2</sub> and other gas pollution that reside in the atmosphere. Many researchers believe that raising of global temperature is connected with several phenomena such as fossil fuel consumption, pesticide use, macroplastic pollution, and the world population. The figure below shows some data collected about phenomena caused by global warming over the years (Figure 4).

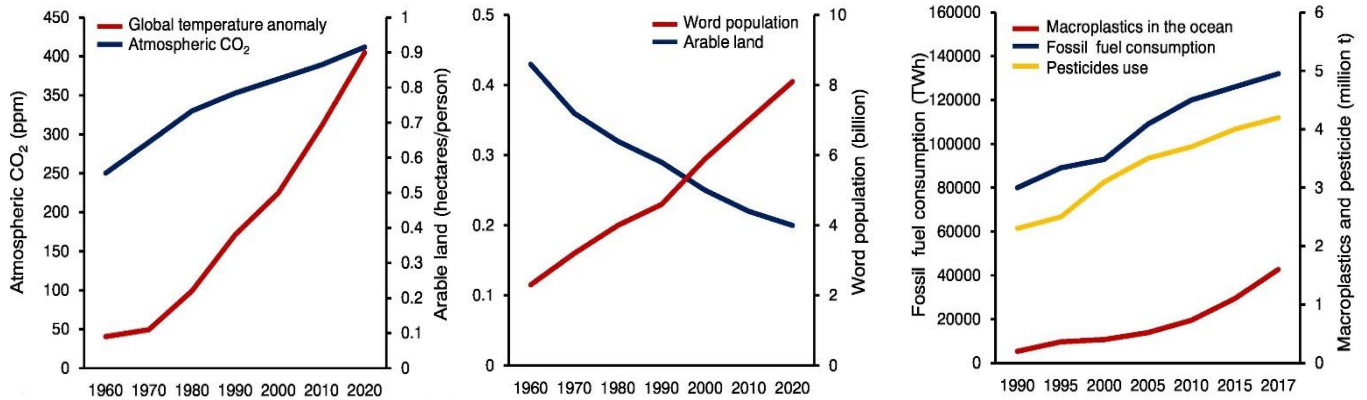


Figure 4. The Global Warming Parameters of the World during 1960-2020 (adapted from Zandalinas et al. (2021))

10. Which of the following statements is correct based on the above data?
- The growing world population correlates with decreasing of microplastic in the oceans
  - The increase of arable land complies with the increasing of global temperature.
  - the increase in fossil fuel consumption is in the same trend with increasing atmospheric CO<sub>2</sub>
  - The decrease in the human population correlates with the increase in the use of pesticides
  - The decrease of atmospheric CO<sub>2</sub> is in the same trend with increasing global temperature;
11. How does fossil fuel consumption affect the raising of global temperature?
- Fuel burning produces carbon dioxide that is trapped in the atmosphere
  - The heat from oil burning will increase the surrounding temperature.
  - The heat from burning fuel will decrease humidity in the air
  - The gas from fuel burning will trap water molecules and decrease rain intensity.
  - the nitrogen gas from oil burning will damage the ozone layer

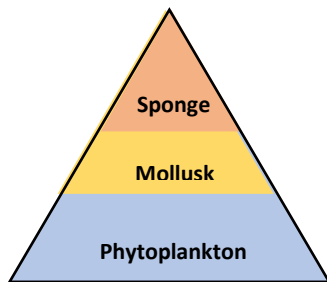
### ***Marine Ecosystem***

Coral reefs are fundamental to the functioning of marine ecosystems and the provision of ecosystem services (Hoegh-Guldberg, 2011). It became the habitat of many organisms such as phytoplankton, coralline algae, zooplankton, sponges, gastropods, molluscs, corals, sea urchins, crabs, fish, and seals. However, in recent decades, coral reefs have been rapidly degraded in response to numerous anthropogenic stressors, including centuries of overfishing, pollution, and global warming. The decline of the wild population in coral reefs disrupts the life cycle and ecosystem balance.

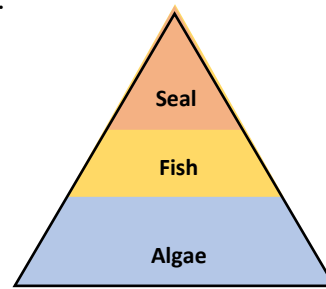
12. Which organism plays a role as producers in coral reef ecosystems?
- sea urchin
  - phytoplankton
  - Gastropods
  - corals
  - sponges

13. Which diagram (ecological pyramid) shows the portrait of the balance ecosystem in the coral reef ecosystem?

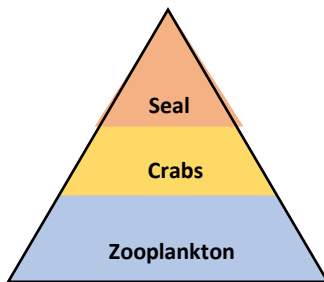
a.



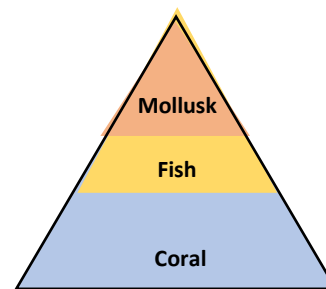
d.



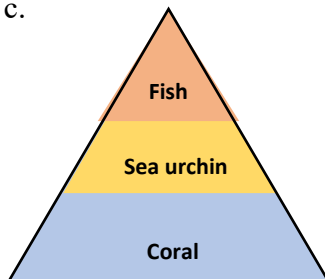
b.



e.



c.



### ***Insect and Pest***

The insect pest *Tribolium confusum* causes great damage in the grain corps. The biological control of predatory bugs, such as *Amphibolus venator* and *Xylocoris flavipes*, has a potential to reduce the pest population because they prey on them. The scientist conducts the research for the effectivity of biological predator in preying insect pest result in Table 3.

Table 3. The biological control of predatory bugs

Combination of treatments	Without <i>T. confusum</i>		With <i>T. confusum</i> (N=10)		
	No. of dead <i>X. flavipes</i>	No. of dead <i>A. venator</i>	No. of dead <i>X. flavipes</i>	No. of dead <i>A. venator</i>	No. of dead <i>T. confusum</i>
4 adults <i>X. flavipes</i> + 2 adults <i>A. venator</i>	3	0	1	0	9
4 adults <i>X. flavipes</i> + 2 nymphs <i>A. venator</i>	3	0	1	0	9
4 nymphs <i>X. flavipes</i> + 2 adults <i>A. venator</i>	1	0	0	0	8
4 nymphs <i>X. flavipes</i> + 2 nymphs <i>A. venator</i>	0	0	0	0	10

14. Based on the result, which predatory bugs have a great impact in reducing the pest population?

- a. *X. flavipes* (nymphs and adult stage)
- b. *X. flavipes* (nymphs) and *A. venator* (nymphs)
- c. *X. flavipes* (nymphs) and *A. venator* (adult)
- d. *X. flavipes* (adult) and *A. venator* (nymphs)
- e. *X. flavipes* (adult) and *A. venator* (adult)

15. What is the interaction between two predators *A. venator* and *X. flavipes* based on the laboratory treatment result?

- a. In the present of *T. confusum*, *A. venator* preys on nymphs *X. flavipes*
- b. In the absence of *T. confusum*, *A. venator* preys on nymphs *X. flavipes*
- c. In the absence of *T. confusum*, *A. venator* preys on adults *X. flavipes*
- d. In absence of *T. confusum*, *X. flavipes* prey on nymphs *A. venator*
- e. In the absence of *T. confusum*, *X. flavipes* preys on adults *A. venator*

References:

*Coral Reefs: An Ecosystem in Transition* (pp. 391–403). Springer Netherlands.

[https://doi.org/10.1007/978-94-007-0114-4\\_22](https://doi.org/10.1007/978-94-007-0114-4_22)

Reece, J. B., Meyers, N., Urry, L. A., Cain, M. L., Wasserman, S. A., & Minorsky, P. V. (2015). *Campbell Biology* (Vol. 10). Pearson Higher Education.

Zandalinas, S. I., Fritschi, F. B., & Mittler, R. (2021). Global Warming, Climate Change, and Environmental Pollution: Recipe for a Multifactorial Stress Combination Disaster. *Trends in Plant Science*, 26(6), 588–599. <https://doi.org/10.1016/j.tplants.2021.02.011>

## **APPENDIX E. Attitude Towards Science Questionnaire**

1. Learning science makes me cheerful
2. Science is relevant to everyday life
3. Science is easy for me
4. I feel helpless when doing science
5. I am sure I can do well on science tests
6. I feel nervous while learning science
7. I want to join a science club
8. I like watching science programs on television/internet
9. I can understand difficult scientific concepts
10. Science is essential for a country's development
11. The benefits of science are greater than the harmful effects
12. I feel enthusiastic while learning science
13. I am not good at science
14. I want to do more science activities outside school
15. I like reading science magazines and books
16. Learning science makes me anxious
17. I feel tense while learning science
18. I cannot understand science even if I try hard
19. Science will help make the world a better place in the future
20. I feel happy while learning science
21. I enjoy learning science
22. I look forward to science lessons in school
23. I feel stressed when I must learn science
24. I like to participate in science class discussion
25. I like to visit science museums
26. Science lesson is exiting
27. I want to study science more deeply than I do at present

## **APPENDIX F. Science Motivation Questionnaire II**

01. The science I learn is relevant to my life.
02. I like to do better than other students on science tests.
03. Learning science is interesting.
04. Getting a good science grade is important to me.
05. I put enough effort into learning science.
06. I use strategies to learn science well.
07. Learning science will help me get a good job.
08. It is important that I get an “A” in science.
09. I am confident I will do well on science tests.
10. Knowing science will give me a career advantage.
11. I spend a lot of time learning science.
12. Learning science makes my life more meaningful.
13. Understanding science will benefit me in my career.
14. I am confident I will do well on science labs and projects.
15. I believe I can master science knowledge and skills.
16. I prepare well for science tests and labs.
17. I am curious about discoveries in science.
18. I believe I can earn a grade of “A” in science.
19. I enjoy learning science.
20. I think about the grade I will get in.
21. I am sure I can understand science.
22. I study hard to learn science.
23. My career will involve science.
24. Scoring high on science tests and labs matters to me.
25. I will use science problem-solving skills in my career.

## ACKNOWLEDGEMENT

I would like to express my gratitude to my supervisor, Dr. Erzsébet Korom for the help and guidance during my PhD journey in the Doctoral School of Education, University of Szeged. We work together to conduct research, complete the thesis, write manuscripts, and other academic activities. I learn a lot about academic life from her. Thanks to her guidance, support, and encouragement, I can pass through difficult times during my studies and be able to give more effort to contribute to academic communities.

I would like to thank to late Prof. Benő Csapó who inspired me to be a good researcher and educator who makes a positive impression on students and the community. Many thanks to Prof. Gyöngyvér Molnár who allowed me to use her instruments and taught me how to be more concise in research and data analysis. Other thanks go to Prof. Tibor Vidákovich and Dr. Edit Katalin Molnar for the valuable lesson during the research seminar. I do realize that I have learned a lot from the others' works as well as the constructive comments and feedback on my research reports. I express special gratitude to all the professors and staff in the Doctoral School of Education for the wonderful lessons and experiences I have gained from the courses and lectures.

I would like to express my thanks to my classmates, seniors, and colleagues from the Doctoral School of Education, University of Szeged and Semarang PGRI University for the moral support and shared experiences throughout my PhD journey. I appreciate all the information and knowledge that I got through our discussions in formal and informal situations. Many thanks to all my Indonesian, Hungarian, and international friends for the many moments and events we have spent together. It fulfills my social life and makes Szeged a nice place to live. My sincere thanks also go to the principal, teachers, and students in Indonesia for their cooperation and participation in my research.

Most importantly, I would like to thank my family for the constant support and unconditional love. Without their encouragement and prayer, I may not have been able to be here and complete my education.

## RELEVANT PUBLICATIONS

### Published papers

- Wicaksono, A. G. C., & Korom, E. (2023). Attitudes towards science in higher education: Validation of questionnaire among science teacher candidates and engineering students in Indonesia. *Heliyon*, 9(9), e20023. <https://doi.org/10.1016/j.heliyon.2023.e20023>
- Wicaksono, A. G. C., & Korom, E. (2023). Role of inductive reasoning, gender, learning satisfaction, and educational and career preference in predicting scientific competency in high school. *Thinking Skills and Creativity*, 49, 101376. <https://doi.org/10.1016/j.tsc.2023.101376>
- Wicaksono, A. G. C., & Korom, E. (2022). Review of problem-solving measurement: an assessment developed in the Indonesian context. *Participatory Educational Research*, 9(1), 116–136. <https://doi.org/10.17275/per.22.7.9.1>

### Conferences

- Wicaksono, A. G. C., & Korom, E. (2023). The profile of high school students' problem-solving ability in Indonesia. *XIX 19th Conference on Educational Assessment : PÉK 2023*, 26. [https://www.edu.u-szeged.hu/pek2023/download/PEK\\_2023\\_CEA\\_2023\\_absztraktkotet.pdf#page=27](https://www.edu.u-szeged.hu/pek2023/download/PEK_2023_CEA_2023_absztraktkotet.pdf#page=27)
- Wicaksono, A. G. C., & Korom, E. (2022). The relationship between science motivation, attitude toward science, and inductive reasoning among science major students in Indonesia. *XVIII 18th Conference on Educational Assessment*, 22.
- Wicaksono, A. G. C., & Korom, E. (2022). The measurement of inductive reasoning skill in higher education in Indonesia. *JURE 2022: Unpredictable Challenges - Education in a Rapidly Changing World*.
- Wicaksono, A. G. C., & Korom, E. (2021). The validation of attitude towards science questionnaire among Indonesian undergraduate students. *EARLI 2021: Education and Citizenship: Learning and Instruction and the Shaping of Futures*, 27