http://journal.lppmunindra.ac.id/index.php/jkpm/
ISSN 2477-2682 (Electronic)
ISSN 2477-2348 (Print)

Jkpm Jurnal Kajian Pendidikan Matematika

Original Research

Development of a Creative Problem-Solving Test based Higher Order Thinking Skills: A Rasch Model Measurement

Farida[1*)], Yosep Aspat Alamsyah[2], & Suherman[3]
[1, 2, 3]Universitas Islam Negeri Raden Intan Lampung, Bandar Lampung, Indonesia
[3]University of Szeged, Szeged, Hungary

INFO ARTICLES

Article History:

Received: 26-04-2025
Revised: 30-06-2025
Approved: 30-06-2025
Publish Online: 30-06-2025

Key Words:

Creative Problem Solving; Rasch Evaluation; Higher Education Student; Higher Order Thinking;

Abstract: Creative Problem-Solving (CPS) is increasingly valued in independent curriculum education, yet comprehensive psychometric validation of CPS assessment tools remains limited. This study aimed to develop and validate an instrument for measuring CPS skills among students using the Rasch Model. A total of 137 higher education students participated. Analysis showed the assessment items fit the Rasch Model well, with strong item-person alignment, effective discrimination among difficulty levels, and good reliability. The Wright map revealed clear patterns between item difficulty and student ability. One item showed Differential Item Functioning (DIF) favoring male students. Additionally, fourth-semester female students outperformed those in later semesters. Significant CPS ability differences were found across gender and between urban and rural students, emphasizing the need for targeted differentiation strategies in education.

Abstrak: Creative Problem-Solving (CPS) semakin dihargai dalam pendidikan kurikulum merdeka, namun validasi psikometrik yang komprehensif terhadap alat asesmen CPS masih terbatas. Penelitian ini bertujuan untuk mengembangkan dan memvalidasi instrumen untuk mengukur keterampilan CPS pada mahasiswa dengan menggunakan Model Rasch. Sebanyak 137 mahasiswa pendidikan tinggi berpartisipasi dalam penelitian ini. Hasil analisis menunjukkan bahwa butir-butir asesmen sesuai dengan Model Rasch, dengan kesesuaian item dan responden yang kuat, diskriminasi tingkat kesulitan yang efektif, dan reliabilitas yang baik. Analisis peta Wright mengungkapkan pola yang jelas antara tingkat kesulitan item dan kemampuan mahasiswa. Satu butir menunjukkan adanya Differential Item Functioning (DIF) yang menguntungkan mahasiswa laki-laki. Selain itu, mahasiswa perempuan semester empat menunjukkan kinerja lebih baik dibandingkan dengan mahasiswa di semester berikutnya. Perbedaan signifikan dalam kemampuan CPS ditemukan berdasarkan gender dan antara mahasiswa dari daerah perkotaan dan pedesaan, menekankan pentingnya strategi diferensiasi yang terarah dalam pendidikan.

Correspondence Address: Jl. Letnan Kolonel H. Endro Suratmin, Sukarame, Kota Bandar Lampung, Lampung, Indonesia, Kode Pos 35131; *e-mail*: farida@radenintan.ac.id

How to Cite (APA 6th Style): Farida, Alamsyah, Y.A., & Suherman. (2025). Development of a Creative Problem-Solving Test based Higher Order Thinking Skills: A Rasch Model Measurement. *JKPM* (*Jurnal Kajian Pendidikan Matematika*), 10(2): 257-274. http://dx.doi.org/10.30998/jkpm.v10i2.28956

Competing Interests Disclosures: The authors declare that they have no significant competing financial, professional or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

INTRODUCTION

In recent years, there has been a growing interest in assessing creative problem-solving (CPS) abilities among students, as it is recognized as a crucial skill in the 21st century (Care & Kim, 2018; Hao et al., 2017). The ability to think creatively and find innovative solutions to complex problems is crucial in a rapidly changing world that demands adaptability and creative thinking (Suherman & Vidákovich, 2022). Understanding and promoting CPS abilities among students are crucial for several reasons. Firstly, fostering creativity equips individuals with the capacity to generate innovative solutions (Lorusso et al., 2021), fostering entrepreneurship (Val et al., 2019), and driving economic growth (Florida, 2014). Additionally, CPS is vital in addressing complex societal challenges, such as sustainability and social inequality (Mitchell & Walinga, 2017).

CPS in an independent curriculum also fosters collaboration and teamwork. By incorporating CPS into the curriculum, students are empowered to approach problems with an open mind and explore multiple perspectives (Burns & Norris, 2009). They are encouraged to question assumptions, challenge conventional wisdom, and seek alternative solutions. This process not only develops their analytical skills but also nurtures their creativity and divergent thinking abilities (Suherman & Vidákovich, 2022). An independent curriculum provides students with the freedom to explore topics of interest and engage in self-directed learning (Lestari et al., 2023). Students are encouraged to work together, leveraging their diverse perspectives and skills to address complex challenges (Utami & Suswanto, 2022). This collaborative environment enhances their interpersonal and communication skills, preparing them for future collaborative endeavors. Furthermore, incorporating CPS into the curriculum prepares students for the demands of the rapidly evolving 21st-century workforce (Stankovic et al., 2017). As the world becomes increasingly complex and interconnected (Steiner, 2009), employers seek individuals who can think critically (Carnevale & Smith, 2013), adapt to change and the work becomes easier (Sousa et al., 2014), and generate innovative solutions (Wolcott et al., 2021).

CPS has gained recognition as a valuable skill set in educational contexts, including independent curriculum. However, its implementation may face several challenges that need to be addressed to ensure its effectiveness and success. Previous research has highlighted the significance of CPS abilities in various educational contexts (Greiff et al., 2013; Wang et al., 2023; Wolcott et al., 2021). One challenge is the lack of teacher training and familiarity with CPS techniques. Studies have shown that educators may struggle to integrate CPS into their teaching practices due to limited knowledge and experience in facilitating CPS activities (van Hooijdonk et al., 2020). This can hinder the effective implementation of CPS and limit its impact on student learning. Moreover, research has explored the factors that contribute to the development of CPS, including the influence of culture (Cho & Lin, 2010), instructional approaches, and individual characteristics (Samson, 2015). However, there is a limited understanding of the specific factors that develop of CPS abilities students.

Regarding assessment, previous studies have explored alternative approaches to evaluate CPS skills. Performance-based assessments, portfolios, and rubrics that assess creativity, critical thinking, metaphorical thinking, and problem-solving abilities have been proposed as more comprehensive and authentic assessment methods (Abosalem, 2016; Farida et al., 2022; Montgomery, 2002; Suastra et al., 2019)a. These approaches provide a more holistic view of students' CPS skills and encourage the development of higher-order thinking abilities. However, research also offers potential solutions, such as teacher professional development, integration within the curriculum, and alternative assessment methods. By reviewing relevant literature, we aim to build upon existing knowledge and identify gaps in understanding, providing a foundation for this study's contribution to the field.

Despite the increasing recognition of the importance of CPS in independent curriculum education, there is a lack of comprehensive psychometric validation for CPS assessment instruments. Validating the measurement instrument is crucial as CPS remains a poorly defined psychological construct from a psychometric perspective (Tang et al., 2020). In the absence of valid and reliable assessments, instructors face challenges in confidently measuring students' CPS learning in the classroom. Therefore, this study aims to validate CPS using the Rasch model by investigating whether the data align with the Rasch model measurement. The reseach questions (RQ) are followed:

1. RQ1: Does the developed instrument demonstrate reliability and validity based on Rasch measurement?
2. RQ2: What are the patterns of interaction between items and persons in the developed instrument based on the Wright map?
3. RQ3: Are there any instrument biases based on gender according to the Differential Item Functioning (DIF) analysis?
4. RQ4: How does students' CPS develop in terms of course grades?

Literature Review

CPS

CPS is a process that enables individuals to apply creative and critical thinking in finding solutions to everyday problems (Eberle & Stanish, 2021). CPS helps to eliminate the tendency to approach problems haphazardly and, as a result, prevents surprises and/or disappointments with the solutions. Students learn to work together or individually to find appropriate and unique solutions to real-world problems they may encounter, using tried and tested methods. Most importantly, they are challenged to think both creatively and critically as they face each problem.

CPS can also be influenced by external factors as an individual's skill in achieving required goals through a creative process to find new solutions. The importance of communication in the educational process means that teachers must also possess various competencies such as personality, communication, social, lifelong learning, methodology, planning, organization, leadership, and assessment, in order to discern the most significant problems for their respondents (Suryanto et al., 2021). Research states that creative problem-solving can provide students with skills to tackle everyday problem-solving (Abdulla Alabbasi et al., 2021). These skills require extensive practice involving the creative process, and these activities are crucial in developing social skills in the field of creativity. Evaluating ideas and involving multiple people in decision-making with creative thinking in everyday life - the process of generating new ideas and still discussing different ways of thinking.

Several researchers have developed instruments to assess CPS abilities. For instance, the study conducted by Hao et al. (2017) focused on the development of a standardized assessment for CPS skills. The researchers recognized the importance of CPS in today's collaborative work environments and aimed to address the practical challenges associated with assessing this complex construct. The study also highlighted the significance of establishing clear scoring rubrics and criteria for evaluating CPS performance. In another study, Harding et al. (2017) focused on measuring CPS using mathematics-based tasks. The study highlighted the potential of mathematics-based tasks for assessing CPS skills. They employed rigorous psychometric analyses to examine the reliability and validity of the assessment instrument.

These instruments developed by different researchers provide valuable resources for assessing CPS abilities. They offer a comprehensive approach to measuring various aspects of CPS, including creative thinking, problem-solving strategies, and collaboration. They offer a comprehensive approach to measuring various aspects of CPS, including creative thinking, problem-solving strategies, and collaboration. Creative thinking is assessed through a child's ability to generate original ideas and explore multiple solutions to a problem (Suherman & Vidákovich, 2024). Problem-solving strategies focus on how children approach challenges, plan steps, and adjust their actions based on outcomes (Ylvén & Granlund, 2015). Collaboration is evaluated by how well children communicate, share responsibilities, listen to others' ideas, and work as a team to achieve a shared goal (Mueller & Fleming, 2001). By using these instruments, researchers and educators can gain insights into individuals' CPS abilities and tailor instructional strategies to enhance students' creative problem-solving skills.

HOTS

The Development of a CPS Test based on HOTS recognizes that HOTS are essential for fostering effective problem-solving abilities and creativity in students. Research has demonstrated a positive correlation between creative thinking skills and problem-solving capabilities, indicating that students who engage in creative thinking are better equipped to navigate complex problems and generate multiple solutions effectively (Farida et al., 2022). Furthermore, the integration of design thinking principles in curricula has been shown to enhance students' abilities to tackle real-world challenges creatively (Guaman-Quintanilla et al., 2023). Methodologies focusing on CPS are pivotal as they promote comprehensive engagement with materials, particularly in mathematics and science education (Khalid et al., 2020). Studies utilizing the Rasch model emphasize the importance of accurately assessing HOTS, highlighting its relevance in educational research aimed at improving learning outcomes (van de Grift et al., 2019) and developing effective assessments in higher education contexts (Abosalem, 2016).

As educational paradigms evolve towards complex cognitive tasks, the significance of HOTS remains critical. Assessments grounded in the Rasch model provide insights into students' cognitive abilities and their application in problem-solving scenarios (Suherman & Vidákovich, 2025b). The ability to solve creative problems necessitates deeper cognitive engagement, characterized by fluency and originality, which can be fostered through targeted educational frameworks (Suherman & Vidákovich, 2022). This emphasis on higher-order skills is essential not only for student success in academic settings but also for future career readiness, as employers increasingly seek individuals capable of innovative thinking and adaptive problem-solving (Ummah & Yuliati, 2020). Therefore, developing instruments that accurately measure HOTS is fundamental to promoting educational practices that prepare students for the complexities of contemporary life (Wilson & Narasuman, 2020).

Rasch Measurement

Rasch measurement is a psychometric approach developed by Georg Rasch in the 1960s (Panayides et al., 2010). It is used to analyze and interpret data from educational and psychological assessments. The Rasch model, also known as the Rasch measurement model or Rasch model for item response theory (IRT), is a mathematical model that relates the probability of a response to an item to the ability or trait level of the individual being assessed (Cappelleri et al., 2014; Rusch et al., 2017).

The Rasch model is based on the principle of probabilistic measurement, which means that it assesses the probability of a particular response pattern given the person's ability and the item's difficulty (Kyngdon, 2008). Individuals with higher abilities should have a higher probability of answering items correctly, reflecting easier difficulty levels (Andrich, 2017). In other words, probabilities are closely related to the differences between item difficulty and individual ability (Boone et al., 2013). The model assumes that the probability of a correct response follows a logistic function and that the item's difficulty and the person's ability can be placed on the same underlying continuum, often referred to as a logit scale. In Rasch measurement, person abilities and item difficulties are calibrated on an interval scale called logits, and the item and person parameters are completely independent (Chan et al., 2021). This means that the measurement of student abilities remains the same regardless of the difficulty level of the items, and item difficulties remain invariant regardless of student abilities or test takers.

Rasch measurement provides several advantages. It allows for the development of linear measures that are independent of the specific items used in the assessment (Caty et al., 2008). This means that scores obtained from different sets of items can be compared and aggregated. Rasch measurement also provides information about the reliability of the measurement and the fit of the data to the model, which helps assess the quality of the assessment instrument. In educational and psychological research, Rasch measurement is commonly used to evaluate the quality of test items, calibrate item difficulty, estimate person ability, and conduct item and person analysis. It has applications in various fields, including educational assessment, health outcomes research, and social sciences (Planinic et al., 2019). By employing the Rasch model, researchers can gain valuable insights into the relationship between individuals and items, refine measurement instruments, and make meaningful inferences about the construct being measured.

METHOD

Participants

In this cross-sectional study, a total of 137 higher education students participated as respondents. The research obtained ethical approval from the Institutional Review Board of the Universitas Islam Negeri Raden Intan Lampung, no.99/2023, which adhered to the institution's ethical standards. These students were selected from the Department of Mathematics Education using a stratified random sampling technique. In this method, the population is first divided into distinct subgroups or "strata" based on specific characteristics—such as year level, academic performance, or gender—to ensure representation from each subgroup. Then, students are randomly selected from each stratum in proportion to their size within the population. This approach increases the likelihood that the sample accurately reflects the diversity of the entire department and reduces sampling bias. This sampling technique was chosen to ensure that the sample population accurately represents the entire population under investigation. The average age of the participants was 20.84 years, with a standard deviation (SD) of 1.34. In terms of gender distribution, 51.8% of the respondents were female, while 48.2% were male. Regarding their residence, the majority of students (50.4%) came from the city, while the remaining 49.6% came from countryside. The characteristics of the respondents are presented in Table 1.

Table 1. The Characteristics of Participants

Characteristics	Frequency	Percentage (%)
Gender		
Female	71	51.8
Male	66	48.2
Grade/Semester		
4	40	29.2
6	35	25.5
8	62	45.3
Place of residence		
City	69	50.4
Countryside	68	49.6
Ethnic		
Lampung	47	34.31
Java	56	40.87
Others	34	24.82

Instruments

The instruments used in this study were developed by researchers and specifically designed to assess the CPS abilities of students. These instruments were aligned with the university curriculum in the higher education to ensure that they effectively measure the desired skills and competencies. A total of 20 items were developed for this purpose, encompassing various aspects of CPS. The items were presented in a multiple-choice format to allow for objective scoring, reduce ambiguity in responses, and enhance reliability across a large sample. Multiple-choice items also enable efficient assessment of diverse problem-solving scenarios while maintaining consistency in how each participant interprets and responds to the tasks. Each item was scored dichotomously, with a score of 1 assigned to a correct response and 0 to an incorrect response, allowing for straightforward analysis of student performance. An example of instrument is in Figure 1.

The items in the instruments were carefully crafted to evaluate students' ability to apply HOTS, critical and creative thinking, problem-solving strategies, and collaboration within the context of real-world challenges. By utilizing these instruments, the researchers aimed to obtain valuable insights into students' CPS abilities and their proficiency in applying these skills in different scenarios. The instruments were developed to provide a reliable and valid measurement of CPS, enabling the researchers to gain a comprehensive understanding of students' strengths and areas for improvement in this domain. The use of these instruments in this study allowed for a systematic and standardized assessment of CPS, providing valuable data that can contribute to enhancing educational practices and curriculum development.

A food entrepreneur produces two types of products: potato chips and crackers. He has the following raw materials available: 500 kg of potato flour, 600 kg of wheat flour, and 400 kg of cooking oil. Each kilogram of potato chips can be sold for Rp10,000 and requires 1 kg of potato flour and 0.5 kg of cooking oil. Each kilogram of crackers can be sold for Rp8,000 and requires 1 kg of wheat flour and 0.25 kg of cooking oil. What is the maximum profit that the entrepreneur can earn?

A. Rp 9,200,000
B. Rp 9,600,000
C. Rp 9,000,000
D. Rp 9,800,000
E. Rp 9,900,000

Figure 1. An Example of CPS-HOTS Test

Procedure

This study involved a one-week data collection period among higher education students to assess their CPS abilities. The CPS test was administered using Google Forms during regular classroom sessions dedicated to the respective courses. Students were provided with access to the test through their laptops or mobile phones and were allotted 90 min to complete it. In the Google Forms survey, students were required to provide their demographic information, including gender, place of residence, ethnic, and grade level. The CPS test consisted of multiple-choice items, designed to evaluate various aspects of CPS skills. Prior to commencing the test, the researcher provided instructions to the students and presented three example items to familiarize them with the question format and characters that would appear in the test. Upon answering all the questions, students submitted their responses by clicking the 'Submit' button, which saved their answers for further analysis.

Data Analysis

The data collected in this study were analyzed using Rasch measurement, a widely used psychometric approach for assessing the fit between the observed data and the underlying measurement model. In this study, we used Winsteps version 4.7.0 (Linacre, 2020) to analyzed the data, the Rasch model was applied to analyze the responses to the CPS test items. The model estimates the difficulty of each item and the ability of each student on a common logit scale. The data obtained from the CPS assessment instrument were analyzed using various Rasch measurement parameters and techniques. The Outfit mean-square (MNSQ) and Outfit z-standardized (ZSTD) were calculated to assess the fit of each item. The Outfit MNSQ indicates the extent to which the observed responses deviate from the expected responses based on the Rasch model; values between 0.5 and 1.5 are generally considered acceptable. The Outfit ZSTD standardizes the MNSQ values to facilitate comparison across items, with values typically expected to fall between -2.0 and +2.0 to indicate adequate fit. The Point-measure correlation (Pt-Measure Corr) was computed to examine the relationship between item difficulty and person ability, where positive correlations close to +1.0 suggest good alignment between items and the underlying construct measured. This correlation coefficient measures the strength of the association between the item responses and the estimated person abilities on the logit scale.

The Wright map, a graphical representation, was utilized to display the distribution of item difficulties and the corresponding abilities of the students. This map provides a comprehensive overview of the item difficulty hierarchy and the range of abilities exhibited by the students. Additionally, a logit value person (LVP) analysis was conducted to identify the CPS abilities students. To explore potential differences in item functioning based on gender and living of residence, DIF analysis was conducted. DIF analysis identifies items that may function differently for different groups, indicating potential bias or differential performance across groups. Additionally, the R package statistics was employed to see the statistical distribution among grades level.

RESULTS

RQ1

The results of the validation analysis using Rasch analysis are presented in Table 2.

Table 2. The outcomes of the Rasch analysis conducted on CPS.

Characteristics	Item	Person
Number items	20	137
infit MNSQ		
Mean	1.00	1.00
SD	0.18	0.14
outfit MNSQ		
Mean	1.02	1.02
SD	0.30	0.29
Separation	4.01	1.40
Reliability	0.66	0.94
Raw variance explained by measures	76.6%	

In this study, we conducted an analysis of the measurement characteristics on the involved items and individuals. A total of 20 items were used in the measurement, and the analysis results showed that the average infit MNSQ for the items was 1.00, with a SD of 0.18. Meanwhile, the average outfit MNSQ for the items was 1.02, with a SD of 0.30. This indicates that the items have a reasonably good fit with the Rasch measurement model criterias. Additionally, the analysis results also showed that the separation for the items was 4.01, indicating the ability to differentiate the difficulty levels among different items. The measurement reliability for the items was obtained at 0.66, indicating a satisfactory level of reliability.

On the other hand, there were a total of 137 individuals who participated in the measurement. The analysis results indicated that the average infit MNSQ for the individuals was 1.00 (SD = 0.14). The average outfit MNSQ for the individuals was 1.02, with a SD of 0.29. The separation of individuals was obtained at 1.40, while the measurement reliability for the individuals was 0.94.

These analysis results provide an overview of the measurement characteristics on the items and individuals in this study, which can serve as a basis for understanding the fit and reliability of the conducted measurement. The internal consistency reliability, as indicated by Cronbach's alpha, was calculated at 0.88, demonstrating good reliability of the instrument. The Standard Error of Measurement (SEM) was estimated to average around 0.25, indicating a reasonably precise measurement of person ability. Regarding unidimensionality, the principal component analysis of residuals revealed that the unexplained variance in the first contrast was approximately 3.4%, well below the commonly accepted cutoff of 5%, thus supporting the instrument's unidimensionality. Finally, the test information function showed that the instrument provided the most precise measurement for person abilities within the range of approximately -2.0 to +2.0 logits, indicating effective discrimination across a broad spectrum of respondent abilities. The distribution of item fit is displayed in Figure 2.

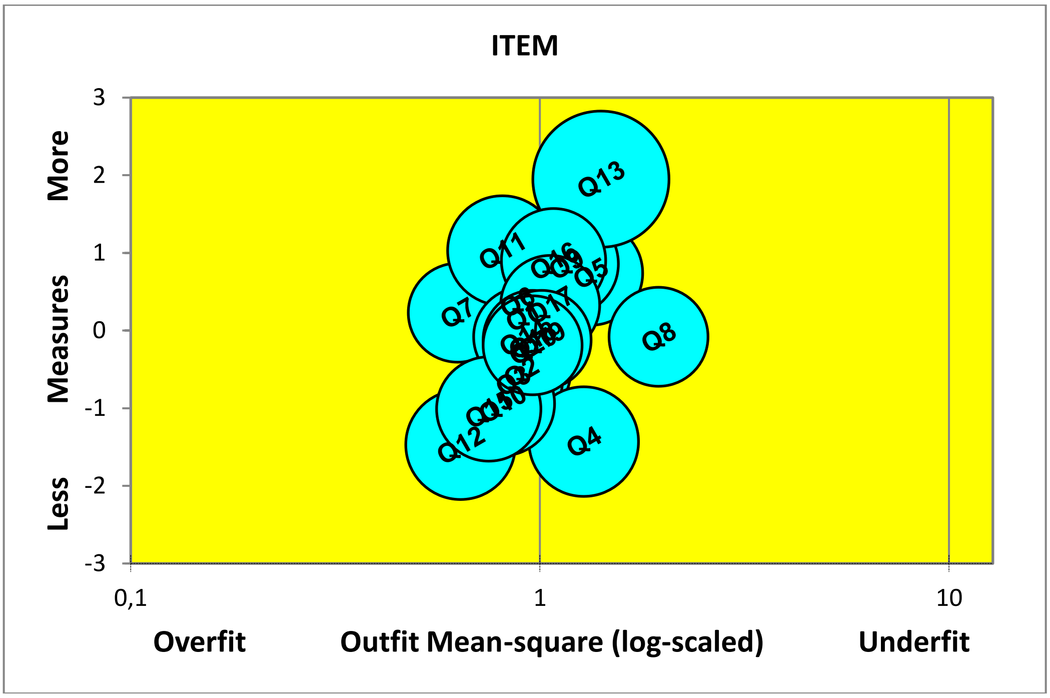


Figure 2. The Distribution Items based on Bubble Map

RQ2

The pattern of interaction between items and individuals in the developed instrument, based on the Wright map, is presented in Figure 3. It can be observed from Figure 3 that the instrument consists of 20 items and involves 137 students as respondents. The vertical line on the right side represents the items, while the left side represents the number of respondents. It can be noted that item number 13 (Q13) falls into the category of easy items, whereas item number 12 (Q12) is categorized as a difficult item. The distribution or characteristics of difficult and easy items can be seen in Figure 4.

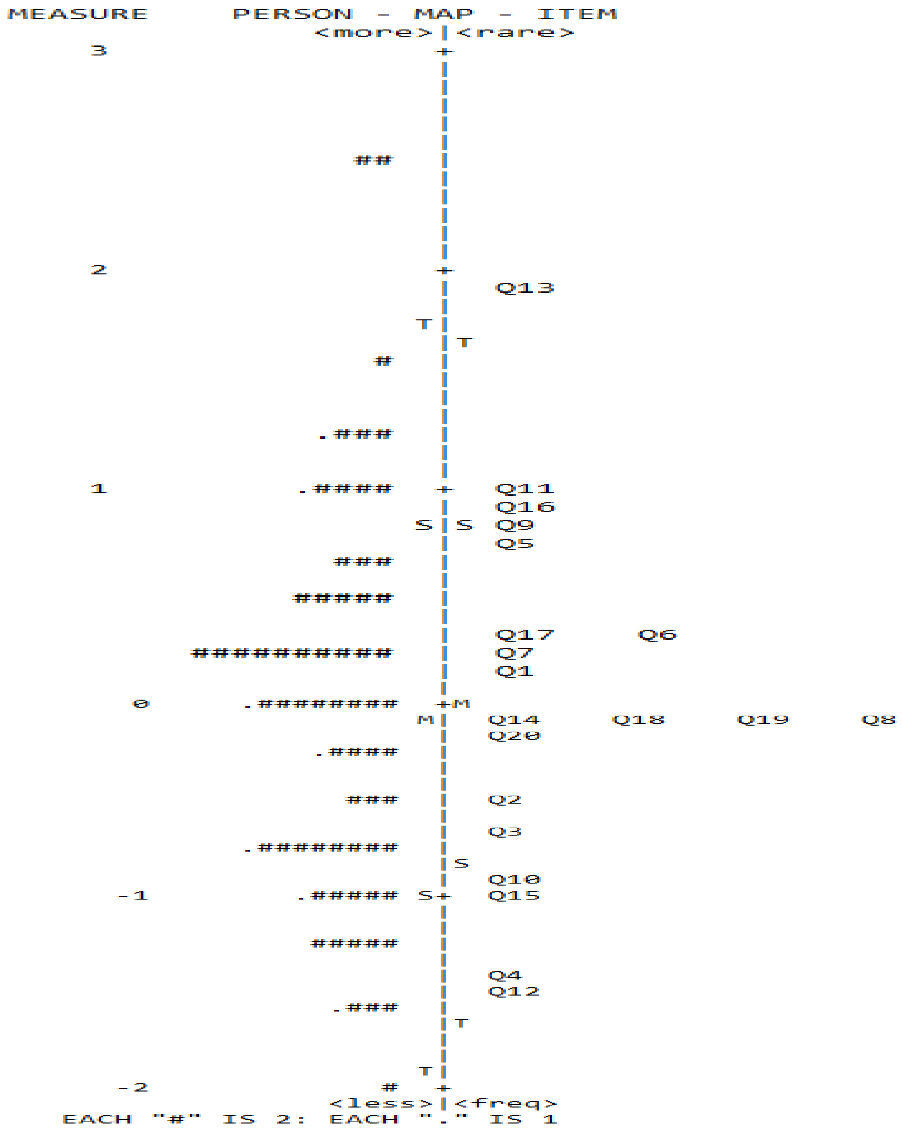
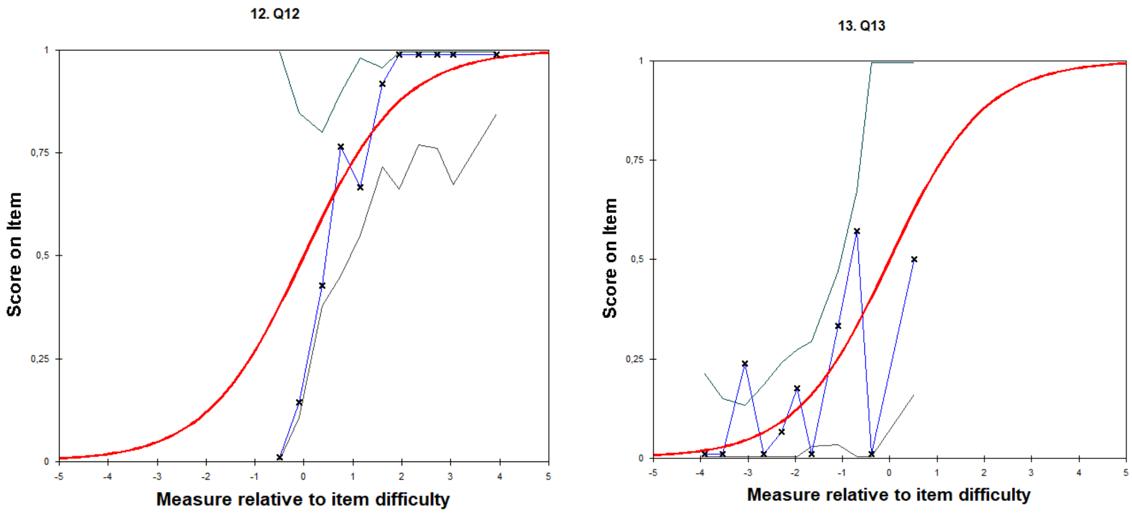


Figure 3. Wright Map



4a. difficult item 4a. easy items

Figure 4. An Item Belongs to Difficult (4a) and Easier Items (4b)

To determine the fit of the developed items based on the Rasch model, three criteria were considered: Outfit MNSQ, Outfit ZSTD, and Pt-Measure Corr. A range between 0.5 and 1.5 for Outfit MNSQ values for both items and individuals indicates good fit between the data and the model. Outfit ZSTD values between -1.9 and 1.9 imply that the items can be predicted. Additionally, Pt-Measure Corr is used to determine if the items measure the intended construct. If the value is positive (+), it indicates that the item measures the intended construct. Conversely, if the value is negative (-), the item does not measure the intended construct.

Table 3. Distribution Item based 3 Criteria of Item Fit.

Items	Measure	Infit MNSQ	Outfit MNSQ	Outfit SZTD	Pt-Measure Cor
8	-0.08	1.60	1.95	8.51	-0.27
13	1.95	1.05	1.41	1.47	0.22
5	0.74	1.11	1.33	2.52	0.22
4	-1.43	1.15	1.28	1.47	0.16
9	0.86	1.09	1.16	1.17	0.27
16	0.90	1.16	1.08	0.63	0.24
17	0.33	1.04	1.06	0.67	0.34
19	-0.12	1.05	1.01	0.19	0.35
20	-0.19	1.02	0.96	-0.41	0.38
18	-0.12	1.00	0.96	-0.43	0.40
6	0.37	0.95	0.88	-1.26	0.45
14	-0.08	0.94	0.91	-1.05	0.45
1	0.19	0.91	0.91	-1.00	0.48
3	-0.64	0.91	0.86	-1.30	0.47
2	-0.53	0.89	0.90	-0.98	0.47
11	1.03	0.89	0.81	-1.35	0.49
10	-0.94	0.86	0.81	-1.48	0.50
15	-1.01	0.82	0.75	-1.90	0.54
12	-1.47	0.80	0.64	-2.13	0.54
7	0.23	0.69	0.63	-4.57	0.70

Based on the three criteria mentioned, it is evident that item 8 (Q8) does not meet the above-mentioned criteria, indicating that the item does not fit well. Therefore, it is recommended to either eliminate or revise item 8.

RQ3

The DIF analysis was conducted to determine if there were any items that favored one gender (in the context of this study). An item is considered to have DIF if the t-value is less than -2.0 or greater than 2.0, the DIF contrast value is less than -0.5 or greater than 0.5, and the $p < 0.05$ or $p > -0.05$ (Bond & Fox, 2015; Boone et al., 2014). Table 4 was the results of the analysis using the Rasch model.

Table 4. Potential DIF Owing Gender

Item	DIF		DIF Contrast	t-value	Prob
	Female	Male			
Q12	-1.97	-1.06	-0.91	-2.04	0.0432

The analysis reveals that item Q12 is a difficult item, indicating that it can differentiate the abilities between males and females. This is further supported by the DIF analysis, which examines the item's performance across gender groups. The DIF graph (Fig.5) provides visual representation of the DIF values for each item.

In the graph, the DIF values for item Q12 are noticeably higher compared to other items. This suggests that there is a significant difference in the performance of males and females on this particular item. The DIF analysis indicates that item Q12 may favor one gender over the other in terms of difficulty or discrimination. These findings are important as they highlight potential gender-related biases in the measurement of collaborative problem-solving abilities. Further investigation and potential item revision may be necessary to ensure fair and unbiased assessment of all individuals, regardless of their gender.

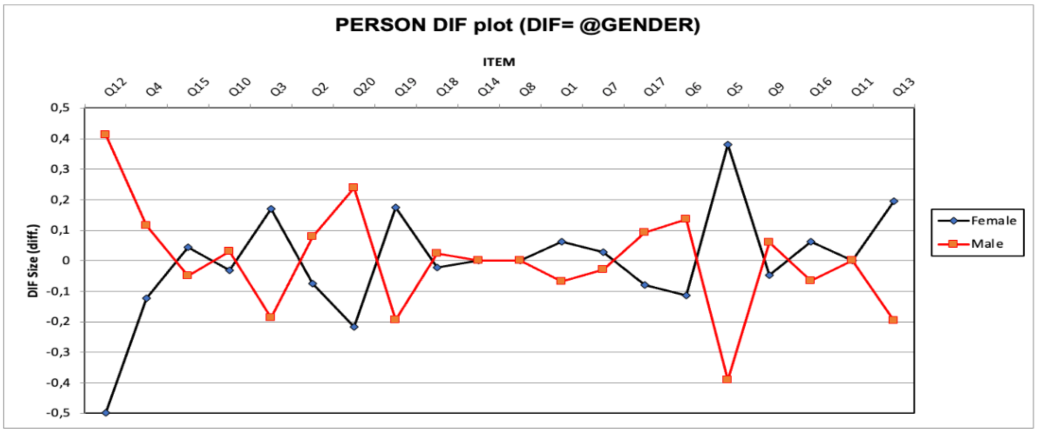
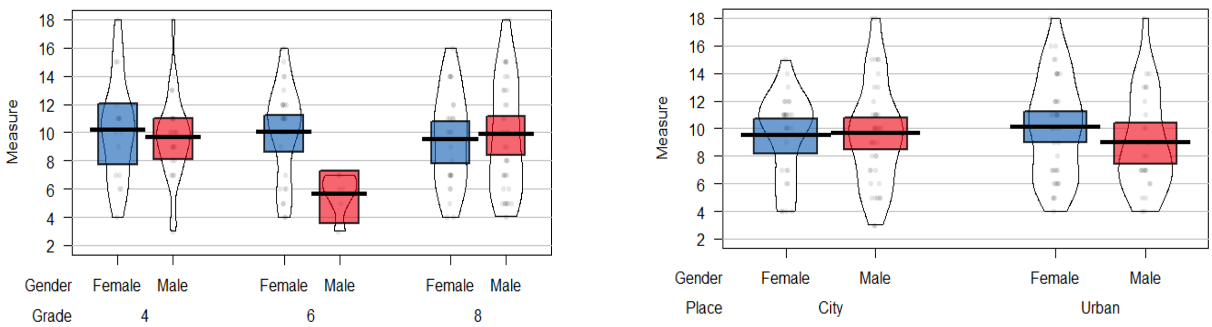


Figure 5. Potential DIF Owing Gender

RQ 4

The statistical description of students' responses to the given items is presented in Figure 5. It can be observed that among female students, those in the 4th semester have an average (M) response ability of M = 10.19, SD = 3.73, followed by those in the 6th semester with M = 10.00, SD = 3.32, and those in the 8th semester with M = 9.50, SD = 3.37. On the other hand, among male students, those in the 4th semester have an average response ability of M = 9.67, SD = 3.11, followed by those in the 6th semester with M = 5.67, SD = 1.51, and those in the 8th semester with M = 9.86, SD = 4.02.

However, when comparing students' abilities based on their place of residence (urban vs. rural), there are differences as shown in Figure 8. Female students residing in urban areas have an average response ability of M = 9.52, SD = 3.11, while those in rural areas have M = 10.10, SD = 3.59. Similarly, male students residing in urban areas have an average response ability of M = 9.68, SD = 3.81, while those in rural areas have M = 9.00, SD = 3.58. These findings indicate that the location of residence may influence students' collaborative problem-solving abilities to some extent.



5a. Grade and gender 5b. Grade and place

Figure 5. The Students' Ability to Answer based on Grade, Gender, and Place of Residence

DISCUSION

Overall, this analysis provides an understanding of the measurement characteristics of the items and individuals in this study. The findings indicate that the measurement used has a reasonably good fit with the Rasch measurement model, the ability to differentiate the difficulty levels among items, and a sufficiently high level of reliability. Support from relevant research in this field also corroborates these findings and provides a strong foundation for understanding the measurement characteristics in this study. Previous studies, such as the one conducted by Komarudin and Suherman (2024), also found similar results in terms of fit, separation, and reliability of the measurement.

The analysis results indicating good fit between items and individuals with the Rasch measurement model serve as an indicator that the measurement accurately represents the measured characteristics. The successful separation of difficulty levels among items also provides an advantage in providing more detailed and accurate information about the measured individual abilities (Farida et al., 2024). This is consistent with previous research stating that adequate separation is crucial to ensure reliable and valid measurement (Suherman & Vidákovich, 2025b).

Additionally, the reasonably good reliability for both items and individuals provides confidence that the obtained measurement results are dependable and consistent. In the context of this study, a reliability of 0.66 for items and 0.94 for individuals indicates a satisfactory level of reliability. Other relevant studies can also provide support for the analysis of measurement characteristics conducted in this study. For example, a study conducted by Chan et al. (2021) found similar results in terms of fit and reliability of the measurement. Overall, this analysis provides a deep understanding of the measurement characteristics of the items and individuals in this study. The support from relevant research and the analysis results showing good fit, separation, and reliability provide confidence that the measurement conducted in this study is reliable and provides valid information about the measured characteristics.

Based on the observed interaction pattern in Figure 4, conclusions can be drawn about the difficulty level of the items. For example, item 13 (Q13) is seen to be positioned lower on the vertical line, indicating that it belongs to the category of easy items. Conversely, item 12 (Q12) is seen to be positioned higher, indicating that it belongs to the category of difficult items (Anderson et al., 2016; Boone et al., 2014).

This analysis provides important information about the difficulty level of each item in the developed instrument. By knowing the difficulty level of the items, adjustments and further development can be made to ensure that the used items cover appropriate levels of difficulty aligned with the research objectives (Chan et al., 2021; Lidinillah et al., 2020; Schneider et al., 2022). However, it should be noted that assessing the difficulty level of the items is not solely based on the position of the items on the vertical line in Figure 3 but also takes into consideration other factors such as the characteristics of the respondents and the deeper context of the measurement.

Moving on to the DIF analysis, DIF refers to differences in response characteristics to an item between two or more groups of respondents who should have the same level of ability (Amalina & Vidákovich, 2023; De Ayala et al., 2002; Fuad et al., 2017). In this context, differences in ability between males and females are explored using the concept of DIF. In this study, it is found that Q12 has the potential to differentiate the ability between males and females, with a DIF category. This indicates that males and females have different probabilities of answering Q12 despite having the same level of ability. In this context, there is an indication that Q12 may be more difficult for one gender group.

However, it is important to note that DIF analysis only provides preliminary indications of potential differences in response characteristics between respondent groups. It is crucial to view these DIF findings as information that can assist in instrument development and gain a better understanding of how the behavioral items in the instrument perform across specific groups. In further research, steps can be taken to evaluate the causes of DIF and ensure that the instrument measures accurately without gender bias.

In our research, we found that female students in the 4th semester demonstrated a higher average ability compared to their male counterparts, indicating potential shifts in academic development that could be influenced by factors like maturity or learning environment (Erdogdu, 2022; Suherman & Vidákovich, 2025a). Additionally, the observed decreases in average ability among female students from the 4th to the 8th semester may suggest challenges faced as they advance in their studies. Research indicates possible relationships between collaborative learning experiences and cumulative learning outcomes (Aluko et al., 2025).

Further analysis regarding residential location reveals that female students from rural areas exhibited higher average abilities compared to their urban counterparts. This trend may arise from differing educational resources and support systems available in these environments (Lin, 2017). The insights gained from comparing male and female students based on urbanity reinforce the necessity of understanding contextual factors that influence educational outcomes. While urban students often have access to modern educational tools, rural students may benefit from traditional learning environments that promote experiential learning (Lin, 2017). The discrepancies in average abilities highlighted in the presented data accentuate the notion that both gender and locality significantly influence students' educational trajectories (Farida et al., 2023).

Moreover, research has established that effective collaborative problem-solving environments can greatly affect students' capacities for creative and critical thinking, highlighting the need for educators to tailor their instructional strategies to address these disparities comprehensively (Supriadi et al., 2024, 2025). Overall, the findings underline the critical nature of utilizing differential teaching methods and assessment structures that accommodate diverse learning needs and environments, enabling all students to reach their potential.

These findings highlight the potential influence of gender, semester, and residential location on the answering abilities of students (Farida et al., 2024; Komarudin & Suherman, 2024; Suherman & Vidakovich, 2022). The variations in means and standard deviations suggest differences in learning experiences, exposure to educational resources, or other contextual factors that may contribute to variations in answering abilities.

Overall, these analyses provide valuable insights into the characteristics of the measurement, differential item functioning, and the relationship between answering abilities and various factors such as gender, semester, and residential location. They contribute to a better understanding of the data and offer implications for future research and instrument development in this field.

Limitation and Future Research

This study provides important contributions to the development of evaluation instruments for assessing CPS students. However, there are some limitations that need to be addressed. Firstly, the measurement reliability for the items obtained a value of 0.66, indicating a moderate level of reliability. While this reliability may be acceptable in some research contexts, improving reliability is desirable for the development of more robust evaluation instruments in the future.

Additionally, there is an item, item 8 (Q8), that does not meet the criteria in the item fit analysis with the Rasch Model. This item could be removed or revised to ensure better fit and validity of the evaluation instrument. Revising and refining items that do not meet the criteria is necessary to ensure that the evaluation instrument produces more accurate and reliable results.

Furthermore, there is potential gender-based DIF observed in item Q12. This indicates instrument bias towards gender in terms of answering difficult questions. In future evaluation instrument development, it is important to address this bias to ensure a more neutral and fair instrument for all participants.

The study also provides information about the interaction patterns between items and individuals in the developed instrument based on the Wright map. By examining these patterns, the difficulty level of each item and the distribution of respondents' abilities can be understood. However, further explanation of the implications of these interaction patterns in evaluation instrument development is not provided.

In addition to these limitations, this study lays a strong foundation for future research in the development of evaluation instruments based on HOTS. Future research can focus on improving instrument reliability, eliminating gender-based instrument bias, and further exploring the interaction patterns between items and individuals.

In future studies, it is important to involve a more representative sample and expand the scope of analysis to obtain more generalizable results. Additionally, instrument validation using other methods that provide additional information on instrument fit, validity, and reliability can also be conducted.

Overall, this study contributes to the development of HOTS-based evaluation instruments using the Rasch Model. Despite some limitations that need to be addressed, this study serves as an important foundation for further research in the development of more effective and robust evaluation instruments.

CONCLUSION

Based on the analysis, the following conclusions can be drawn: Measurement using the Rasch model demonstrates a good fit between the items and individuals in the evaluation instrument. The items exhibit good fit with the Rasch model, allowing for differentiation of difficulty levels among different items, and they also have a reasonably good level of reliability. There is an interaction pattern between items and individuals in the evaluation instrument based on the Wright map. The items in the instrument can effectively differentiate the abilities of individuals, with some items being relatively easy and others being more challenging. There are items that do not meet the DIF criteria based on gender. Item Q12 in the evaluation instrument tends to favor males over females in terms of answering ability. Female students in the fourth semester have higher average answering abilities compared to female students in the sixth and eighth semesters. However, there are differences in answering abilities between male and female students, as well as between students living in urban and rural areas.

These conclusions indicate that the development of CPS evaluation instruments can yield valid and reliable measurement results. However, it should be noted that there are some items that need to be improved for greater accuracy. Additionally, there is an indication of instrument bias based on gender in terms of answering abilities. This should be considered when developing instruments that are more gender-neutral and fair in measuring participants' abilities.

In the development of CPS evaluation instruments and the use of the Rasch model, there are positive impacts on curriculum and instruction development. It is important for educators to adopt effective differentiation approaches, ensure gender-neutral evaluation instruments, and consider contextual factors in designing inclusive and learner-centered instruction that aligns with students' potential. Thus, education can become more relevant, responsive, and empower learners to face the challenges of a complex world.

ACKNOWLEDGEMENT

We are thankful to the Institute for Research and Community Service, Universitas Islam Negeri Raden Intan Lampung, Indonesia, for their research support. We also thank the Doctoral School of Education, University of Szeged, Hungary, which provided suggestions and discussions for an idea in the research view.

REFERENCES

Abdulla Alabbasi, A. M., Hafsyan, A. S., Runco, M. A., & AlSaleh, A. (2021). Problem finding, divergent thinking, and evaluative thinking among gifted and nongifted students. *Journal for the Education of the Gifted*, *44*(4), 398–413. https://doi.org/10.1515/ctra-2018-0019

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

Aluko, H. A., Aluko, A., Offiah, G. A., Ogunjimi, F., Aluko, A. O., Alalade, F. M., Ogeze Ukeje, I., & Nwani, C. H. (2025). Exploring the effectiveness of AI-generated learning materials in facilitating active learning strategies and knowledge retention in higher education. *International Journal of Organizational Analysis*. https://doi.org/10.1108/IJOA-07-2024-4632

Amalina, I. K., & Vidákovich, T. (2023). Development and differences in mathematical problem-solving skills: A cross-sectional study of differences in demographic backgrounds. *Heliyon*, *9*(5). https://doi.org/10.1016/j.heliyon.2023.e16366

Anderson, L. M., Reilly, E. E., Gorrell, S., Schaumberg, K., & Anderson, D. A. (2016). Gender-based differential item function for the difficulties in emotion regulation scale. *Personality and Individual Differences*, *92*, 87–91. https://doi.org/10.1016/j.paid.2015.12.016

Andrich, D. (2017). Advances in social measurement: A Rasch measurement theory. In *Perceived health and adaptation in chronic disease* (pp. 66–91). Routledge.

Bond, T., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Third Edition* (3rd edn). Routledge. https://doi.org/10.4324/9781315814698

Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Burns, D. P., & Norris, S. P. (2009). Open-minded environmental education in the science classroom. *Paideusis*, *18*(1), 36–43. https://doi.org/10.7202/1072337ar

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, *36*(5), 648–662. https://doi.org/10.1016/j.clinthera.2014.04.006

Care, E., & Kim, H. (2018). Assessment of twenty-first century skills: The issue of authenticity. *Assessment and Teaching of 21st Century Skills: Research and Applications*, 21–39. https://doi.org/10.1007/978-3-319-65368-6_2

Carnevale, A. P., & Smith, N. (2013). Workplace basics: The skills employees need and employers want. In *Human Resource Development International* (Vol. 16, Issue 5, pp. 491–501). Taylor & Francis. https://doi.org/10.1080/13678868.2013.821267

Caty, G. D., Arnould, C., Stoquart, G. G., Thonnard, J.-L., & Lejeune, T. M. (2008). ABILOCO: a Rasch-built 13-item questionnaire to assess locomotion ability in stroke patients. *Archives of Physical Medicine and Rehabilitation*, *89*(2), 284–290. https://doi.org/10.1016/j.apmr.2007.08.155

Chan, S. W., Looi, C. K., & Sumintono, B. (2021). Assessing computational thinking abilities among Singapore secondary students: A rasch model measurement analysis. *Journal of Computers in Education*, *8*(2), 213–236. https://doi.org/10.1007/s40692-020-00177-2

Cho, S., & Lin, C.-Y. (2010). Influence of family processes, motivation, and beliefs about intelligence on creative problem solving of scientifically talented individuals. *Roeper Review*, *33*(1), 46–58. https://doi.org/10.1080/02783193.2011.530206

De Ayala, R. J., Kim, S.-H., Stapleton, L. M., & Dayton, C. M. (2002). Differential Item Functioning: A Mixture Distribution Conceptualization. *International Journal of Testing*, *2*(3–4), 243–276. https://doi.org/10.1080/15305058.2002.9669495

Eberle, B., & Stanish, B. (2021). *CPS for kids: A resource book for teaching creative problem-solving to children*. Routledge.

Erdogdu, F. (2022). ICT, learning environment and student characteristics as potential cross-country predictors of academic achievement. *Education and Information Technologies*, *27*(5), 7135–7159. https://doi.org/10.1007/s10639-021-10848-x

Farida, F., Alamsyah, Y. A., Anggoro, B. S., Andari, T., & Lusiana, R. (2024). Rasch Measurement Validation of an Assessment Tool for Measuring Students' Creative Problem-Solving through the Use of ICT. *Pixel-Bit. Revista de Medios y Educación*, *71*, 83–106. https://doi.org/10.12795/pixelbit.107973

Farida, F., Alamsyah, Y. A., & Suherman, S. (2023). Assessment in educational context: The case of environmental literacy, digital literacy, and its relation to mathematical thinking skill. *Revista de Educación a Distancia (RED)*, *23*(76), 1–26. https://doi.org/10.6018/red.552231

Farida, F., Supriadi, N., Andriani, S., Pratiwi, D. D., Suherman, S., & Muhammad, R. R. (2022). STEM approach and computer science impact the metaphorical thinking of Indonesian students'. *Revista de Educación a Distancia (RED)*, *22*(69). https://doi.org/10.6018/red.493721

Florida, R. (2014). The creative class and economic development. *Economic Development Quarterly*, *28*(3), 196–205. https://doi.org/10.1177/0891242414541693

Fuad, N. M., Zubaidah, S., Mahanal, S., & Suarsini, E. (2017). Improving Junior High Schools' Critical Thinking Skills Based on Test Three Different Models of Learning. *International Journal of Instruction*, *10*(1), 101–116.

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

Guaman-Quintanilla, S., Everaert, P., Chiluiza, K., & Valcke, M. (2023). Impact of design thinking in higher education: A multi-actor perspective on problem solving and creativity. *International Journal of Technology and Design Education*, *33*(1), 217–240. https://doi.org/10.1007/s10798-021-09724-z

Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. C. (2017). Initial steps towards a standardized assessment for collaborative problem solving (CPS): Practical challenges and strategies. *Innovative Assessment of Collaboration*, 135–156. https://doi.org/10.1007/978-3-319-33261-1_9

Harding, S.-M. E., Griffin, P. E., Awwal, N., Alom, B. M., & Scoular, C. (2017). Measuring collaborative problem solving using mathematics-based tasks. *AERA Open*, *3*(3), 2332858417728046. https://doi.org/10.1177/2332858417728046

Khalid, M., Saad, S., Hamid, S. R. A., Abdullah, M. R., Ibrahim, H., & Shahrill, M. (2020). Enhancing creativity and problem solving skills through creative problem solving in teaching mathematics. *Creativity Studies*, *13*(2), 270–291. https://doi.org/10.3846/cs.2020.11027

Komarudin, K., & Suherman, S. (2024). An Assessment of Technological Pedagogical Content Knowledge (TPACK) among Pre-service Teachers: A Rasch Model Measurement [Evaluación del conocimiento tecnológico pedagógico del contenido (TPACK) entre los profesores en formación: modelo de medición Rasch]. *Pixel-Bit. Revista de Medios y Educación*, *71*, 59–82. https://doi.org/10.12795/pixelbit.107599

Kyngdon, A. (2008). The Rasch model from the perspective of the representational theory of measurement. *Theory & Psychology*, *18*(1), 89–109. https://doi.org/10.1177/095935430708692

Lestari, W., Sari, M. M., Istyadji, M., & Fahmi, F. (2023). Analysis of Implementation of the Independent Curriculum in Science Learning at SMP Negeri 1 Tanah Grogot Kalimantan Timur, Indonesia. *Repository Universitas Lambung Mangkurat*. https://doi.org/10.36348/jaep.2023.v07i06.001

Lidinillah, D. A. M., Aprilia, M., Suryana, D., & Ahmad, A. B. (2020). Development of Creativity Instrument through Rasch Model Analysis. *Universal Journal of Educational Research*, *8*(4), 1620–1627. https://doi.org/10.13189/ujer.2020.080455

Lin, C.-Y. (2017). Threshold effects of creative problem-solving attributes on creativity in the math abilities of taiwanese upper elementary students. *Education Research International*, *2017*(1), 4571383. https://doi.org/10.1155/2017/4571383

Linacre, J. M. (2020). *Winsteps®(Version 4.7. 0)[Computer Software].(4.7. 0). Winsteps. Com*.

Lorusso, L., Lee, J. H., & Worden, E. A. (2021). Design thinking for healthcare: Transliterating the creative problem-solving method into architectural practice. *HERD: Health Environments Research & Design Journal*, *14*(2), 16–29. https://doi.org/10.1177/193758672199422

Mitchell, I. K., & Walinga, J. (2017). The creative imperative: The role of creativity, creative problem solving and insight as key drivers for sustainability. *Journal of Cleaner Production*, *140*, 1872–1884. https://doi.org/10.1016/j.jclepro.2016.09.162

Montgomery, K. (2002). Authentic tasks and rubrics: Going beyond traditional assessments in college teaching. *College Teaching*, *50*(1), 34–40. https://doi.org/10.1080/87567550209595870

Mueller, A., & Fleming, T. (2001). Cooperative learning: Listening to how children work at school. *The Journal of Educational Research*, *94*(5), 259–265. https://doi.org/10.1080/00220670109598761

Panayides, P., Robinson, C., & Tymms, P. (2010). The assessment revolution that has passed England by: Rasch measurement. *British Educational Research Journal*, *36*(4), 611–626. https://doi.org/10.1080/01411920903018182

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 020111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Rusch, T., Lowry, P. B., Mair, P., & Treiblmaier, H. (2017). Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management*, *54*(2), 189–203. https://doi.org/10.1016/j.im.2016.06.005

Samson, P. L. (2015). Fostering student engagement: Creative problem-solving in small group facilitations. *Collected Essays on Learning and Teaching*, *8*, 153–164. https://doi.org/10.22329/celt.v8i0.4227

Schneider, M. C., Agrimson, J., & Veazey, M. (2022). The relationship between item developer alignment of items to range achievement-level descriptors and item difficulty: Implications for validating intended score interpretations. *Educational Measurement: Issues and Practice*, *41*(2), 12–24. https://doi.org/10.1111/emip.12480

Sousa, F. C., Monteiro, I. P., Walton, A. P., & Pissarra, J. (2014). Adapting creative problem solving to an organizational context: A study of its effectiveness with a student population. *Creativity and Innovation Management*, *23*(2), 111–120. https://doi.org/10.1111/caim.12070

Stankovic, J. A., Sturges, J. W., & Eisenberg, J. (2017). A 21st century cyber-physical systems education. *Computer*, *50*(12), 82–85. https://doi.org/10.1109/MC.2017.4451222

Steiner, G. (2009). The Concept of Open Creativity: Collaborative Creative Problem Solving for Innovation Generation-a Systems Approach. *Journal of Business & Management*, *15*(1).

Suastra, I. W., Ristiati, N. P., Adnyana, P. P. B., & Kanca, N. (2019). The effectiveness of Problem Based Learning-Physics module with authentic assessment for enhancing senior high school students' physics problem solving ability and critical thinking ability. *Journal of Physics: Conference Series*, *1171*(1), 012027. https://doi.org/10.1088/1742-6596/1171/1/012027

Suherman, S., & Vidákovich, T. (2022). Assessment of Mathematical Creative Thinking: A Systematic Review. *Thinking Skills and Creativity*, 101019. https://doi.org/10.1016/j.tsc.2022.101019

Suherman, S., & Vidakovich, T. (2022). Tapis patterns in the context of cthnomathematics to assess students' creative thinking in mathematics: A rasch measurement. *Mathematics Teaching Research Journal*, *14*(4), 56–72.

Suherman, S., & Vidákovich, T. (2024). Mathematical creative thinking-ethnomathematics based test: Role of attitude toward mathematics, creative style, ethnic identity, and parents' educational level. *Revista de Educación a Distancia (RED)*, *24*(77). https://doi.org/10.6018/red.581221

Suherman, S., & Vidákovich, T. (2025a). Creative self-efficacy, attitudes, creative style, and environmental literacy: Promoting mathematical creative thinking. *The Journal of Educational Research*, 1–11. https://doi.org/10.1080/00220671.2025.2495329

Suherman, S., & Vidákovich, T. (2025b). Ethnomathematical test for mathematical creative thinking. *Journal of Creativity*, *35*(2), 100099. https://doi.org/10.1016/j.yjoc.2025.100099

Supriadi, N., Jamaluddin, W., Suherman, S., & Komarudin, K. (2025). The Role of Blended Learning in Improving Students' Numerical Ability and Learning Creativity. *Revista de Educación a Distancia (RED)*, *25*(81). https://doi.org/10.6018/red.619061

Supriadi, N., Jamaluddin Z, W., & Suherman, S. (2024). The role of learning anxiety and mathematical reasoning as predictor of promoting learning motivation: The mediating role of mathematical problem solving. *Thinking Skills and Creativity*, *52*, 101497. https://doi.org/10.1016/j.tsc.2024.101497

Suryanto, H., Degeng, I. N. S., Djatmika, E. T., & Kuswandi, D. (2021). The effect of creative problem solving with the intervention social skills on the performance of creative tasks. *Creativity Studies*, *14*(2), 323–335. https://doi.org/10.3846/cs.2021.12364

Tang, T., Vezzani, V., & Eriksson, V. (2020). Developing critical thinking, collective creativity skills and problem solving through playful design jams. *Thinking Skills and Creativity*, *37*, 100696. https://doi.org/10.1016/j.tsc.2020.100696

Ummah, I. K., & Yuliati, N. (2020). The Effect of Jumping Task Based on Creative Problem Solving on Students' Problem Solving Ability. *International Journal of Instruction*, *13*(1), 387–406. https://doi.org/10.29333/iji.2020.13126a

Utami, Y. P., & Suswanto, B. (2022). The Educational Curriculum Reform in Indonesia: Supporting "Independent Learning Independent Campus (MBKM)". *SHS Web of Conferences*, *149*. https://doi.org/10.1051/shsconf/202214901041

Val, E., Gonzalez, I., Lauroba, N., & Beitia, A. (2019). How can Design Thinking promote entrepreneurship in young people? *The Design Journal*, *22*(sup1), 111–121. https://doi.org/10.1080/14606925.2019.1595853

van de Grift, W. J., Houtveen, T. A., Van den Hurk, H. T., & Terpstra, O. (2019). Measuring teaching skills in elementary education using the Rasch model. *School Effectiveness and School Improvement*, *30*(4), 455–486. https://doi.org/10.1080/09243453.2019.1577743

van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & van Tartwijk, J. (2020). Creative problem solving in primary education: Exploring the role of fact finding, problem finding, and solution finding across tasks. *Thinking Skills and Creativity*, *37*, 100665. https://doi.org/10.1016/j.tsc.2020.100665

Wang, M., Yu, R., & Hu, J. (2023). The relationship between social media-related factors and student collaborative problem-solving achievement: An HLM analysis of 37 countries. *Education and Information Technologies*, 1–19. https://doi.org/10.1007/s10639-023-11763-z

Wilson, D. M., & Narasuman, S. (2020). Investigating Teachers' Implementation and Strategies on Higher Order Thinking Skills in School Based Assessment Instruments. *Asian Journal of University Education*, *16*(1), 70–84. https://doi.org/10.24191/ajue.v16i1.8991

Wolcott, M. D., McLaughlin, J. E., Hubbard, D. K., Rider, T. R., & Umstead, K. (2021). Twelve tips to stimulate creative problem-solving with design thinking. *Medical Teacher*, *43*(5), 501–508. https://doi.org/10.1080/0142159X.2020.1807483

Ylvén, R., & Granlund, M. (2015). Collaborative problem solving in the context of early childhood intervention–the link between problems and goals. *Scandinavian Journal of Disability Research*, *17*(3), 221–239. https://doi.org/10.1080/15017419.2013.859175