PIXEL BIT

Nº 71 SEPTIEMBRE 2024
MONOGRÁFICO

e-ISSN:2171-7966
ISSN:1133-8482

Revista de Medios y Educación

PIXEL-BIT

REVISTA DE MEDIOS Y EDUCACIÓN

Nº 71 - SEPTIEMBRE- 2024

https://revistapixelbit.com

Píxel-Bit. Revista de Medios y Educación. 2024 - ISSN: 1133-8482. e-ISSN: 2171-7966.

EDITORIAL
UNIVERSIDAD DE SEVILLA

Ciencias de la Educación
UNIVERSIDAD D SEVILLA

PIXEL-BIT
REVISTA DE MEDIOS Y EDUCACIÓN
SEPTIEMBRE 2024

Nº 71

EQUIPO EDITORIAL (EDITORIAL BOARD)

Píxel-Bit. Revista de Medios y Educación. 2023 - ISSN: 1133-8482. e-ISSN: 2171-7966.

Píxel-Bit. Revista de Medios y Educación. 2023 - ISSN: 1133-8482. e-ISSN: 2171-7966.

FACTOR DE IMPACTO (IMPACT FACTOR)

SCOPUS Q1 Education: Posición 236 de 1406 (83% Percentil). CiteScore Tracker 2022: 5,6 - Journal Citation Indicator (JCI). Emerging Sources Citation Index (ESCI). Categoria: Education & Educational Research. Posición 257 de 739. Cuartil Q2 (Percentil: 65.29) - FECYT: Ciencias de la Educación. Cuartil 1. Posición 16. Puntuación: 35,68- DIALNET MÉTRICAS (Factor impacto 2021: 1.72. Q1 Educación. Posición 12 de 228) - REDIB Calificación Glogal: 29,102 (71/1.119) Percentil del Factor de Impacto Normalizado: 95,455- ERIH PLUS - Clasificación CIRC: B- Categoría ANEP: B - CARHUS (+2018): B - MIAR (ICDS 2020): 9,9 - Google Scholar (global): h5: 42; Mediana: 42 - Journal Scholar Metric Q2 Educación. Actualización 2016 Posición: 405a de 1,115- Criterios ANECA: 20 de 21 - INDEX COPERNICUS Puntuación ICV 2019: 95.10

Píxel-Bit, Revista de Medios y Educación está indexada entre otras bases en: SCOPUS, Fecyt, DOAJ, Iresie, ISOC (CSIC/CINDOC), DICE, MIAR, IN-RECS, RESH, Ulrich's Periodicals, Catálogo Latindex, Biné-EDUSOL, Dialnet, Redinet, OEI, DOCE, Scribd, Redalyc, Red Iberoamericana de Revistas de Comunicación y Cultura, Gage Cengage Learning, Centro de Documentación del Observatorio de la Infancia en Andalucía. Además de estar presente en portales especializados, Buscadores Científicos y Catálogos de Bibliotecas de reconocido prestigio, y pendiente de evaluación en otras bases de datos.

EDITA (PUBLISHED BY)

Grupo de Investigación Didáctica (HUM-390). Universidad de Sevilla (España). Facultad de Ciencias de la Educación. Departamento de Didáctica y Organización Educativa. C/ Pirotecnia s/n, 41013 Sevilla.
Dirección de correo electrónico: revistapixelbit@us.es . URL: https://revistapixelbit.com/
ISSN: 1133-8482; e-ISSN: 2171-7966; Depósito Legal: SE-1725-02
Formato de la revista: 16,5 x 23,0 cm

PIXEL-BIT
REVISTA DE MEDIOS Y EDUCACIÓN

Índice

Nº 70
MONOGRÁFICO
SEPTIEMBRE 2024

Píxel-Bit. Revista de Medios y Educación. 2024 - ISSN: 1133-8482. e-ISSN: 2171-7966.

PIXEL-BIT
e-ISSN: 2171-7966
REVISTA DE MEDIOS Y EDUCACIÓN
MEDIA & EDUCATION JOURNAL

Rasch Measurement Validation of an Assessment Tool for Measuring Students' Creative Problem-Solving through the Use of ICT

Validación de la Medición Rasch de una Herramienta de Evaluación para Medir la Resolución Creativa de Problemas de los Estudiantes a través del Uso de las TIC

Farida Farida
Universitas Islam Negeri Raden Intan Lampung. Indonesia

Yosep Aspat Alamsyah
Universitas Islam Negeri Raden Intan Lampung. Indonesia

Bambang Sri Anggoro
Universitas Islam Negeri Raden Intan Lampung. Indonesia

Tri Andari
Universitas PGRI Madiun. Indonesia

Restu Lusiana
Universitas PGRI Madiun. Indonesia

Received: 2024/04/03; **Revised:** 2024/05/14; **Accepted:** 2024/07/10; **Online First:** 2024/07/19 ; **Published:** 2024/09/01

ABSTRACT
Despite increasing recognition of the importance of creative problem solving (CPS) through the use of ICT in independent curriculum education, there is a lack of comprehensive psychometric validation for CPS assessment instruments. This study aimed to develop and evaluate an assessment instrument to measure CPS through the use of ICT students using the Rasch model. A total of 137 higher education students participated as respondents. For this purpose, 20 items were created, covering different aspects of CPS. Data analysis was performed using Winstep and SPSS software. The Rasch model was employed to confirm the validity and reliability of the newly developed measurement instrument. The findings of the analysis of the Rasch model indicated a good fit between the assessment items and individual students. The items demonstrated adequate fit with the Rasch model, allowing for differentiation of difficulty levels among different items and exhibiting a satisfactory level of reliability. The Wright map analysis revealed patterns of interaction between the items and individuals, effectively discriminating between varying levels of student abilities. In particular, an item showed DIF based on gender, which favours male students in terms of their response abilities. Furthermore, the study identified that female students in the fourth semester exhibited higher average response abilities compared to female students in the sixth and eighth semesters. Furthermore, significant differences in response abilities were observed between male and female students, as well as between students who resides in urban and rural areas. These findings are crucial for educators, emphasising the need to implement effective differentiation strategies.

RESUMEN
A pesar del creciente reconocimiento de la importancia de la resolución creativa de problemas (CPS) a través del uso de las TIC en la educación con un currículo independiente, existe una falta de validación psicométrica integral para los instrumentos de evaluación de CPS. Este estudio tuvo como objetivo desarrollar y evaluar un instrumento de evaluación para medir la CPS a través del uso de las TIC en estudiantes, utilizando el modelo Rasch. Participaron un total de 137 estudiantes de educación superior como encuestados. Para este propósito, se crearon 20 ítems que cubrían diferentes aspectos de la CPS. El análisis de datos se realizó utilizando el software Winstep y SPSS. Se empleó el modelo Rasch para confirmar la validez y fiabilidad del instrumento de medición recién desarrollado. Los hallazgos del análisis del modelo Rasch indicaron un buen ajuste entre los ítems de evaluación y los estudiantes individuales. Los ítems demostraron un ajuste adecuado con el modelo Rasch, lo que permitió diferenciar los niveles de dificultad entre diferentes ítems y mostró un nivel satisfactorio de fiabilidad. El análisis del mapa de Wright reveló patrones de interacción entre los ítems y los individuos, discriminando efectivamente entre los diversos niveles de habilidades de los estudiantes. En particular, un ítem mostró DIF basado en el género, lo que favorece a los estudiantes varones en términos de sus habilidades de respuesta. Además, el estudio identificó que las estudiantes en el cuarto semestre exhibieron habilidades de respuesta promedio más altas en comparación con las estudiantes en el sexto y octavo semestre. Además, se observaron diferencias significativas en las habilidades de respuesta entre estudiantes varones y mujeres, así como entre estudiantes que residen en áreas urbanas y rurales. Estos hallazgos son cruciales para los educadores, enfatizando la necesidad de implementar estrategias de diferenciación efectivas.

KEYWORDS · PALABRAS CLAVES
Creative problem solving, the use of ICT, education, gender, psychometric, Rasch measurement.
Resolución creativa de problemas, uso de TIC, educación, género, psicometría, medición Rasch.

1. Introduction

In recent years, there has been a growing interest in evaluating creative problem solving (CPS) skills among students using information and communication technologies (ICT), as it is recognised 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). Furthermore, ICT plays a central role within the DigCompEdu framework, where technologies are integrated into teaching practices in a pedagogically meaningful way (Caena & Redecker, 2019). Understanding and promoting CPS abilities among students is crucial for several reasons. Firstly, fostering creativity equips people with the capacity to generate innovative solutions (Lorusso et al., 2021), foster entrepreneurship (Val et al., 2019), and drive economic growth (Florida, 2014). Additionally, CPS is vital in addressing complex societal challenges, such as sustainability and social inequality (Mitchell & Walinga, 2017).

As an implication of the 21st century era, numerous nations have acknowledged the necessity of incorporating abilities such as creatice problem-solving (D. Lee & Lee, 2024), computational thinking (Küçükaydın et al., 2024), ICT (Rahimi & Oh, 2024), and creativity (Suherman & Vidákovich, 2024), which are identified as essential skills for the 21st century. These competencies are increasingly being integrated into educational curricula to prepare students for the demands of modern society and the evolving job market (Abina et al., 2024). As such, the emphasis on developing these skills reflects a global recognition of their importance for future success (Yu & Duchin, 2024). The impact of the 21st century era extends beyond education into the workforce and daily life. Research by Arredondo-Trapero et al. (2024) emphasizes that problem-solving, critical thinking, and ICT are crucial for innovation and competitiveness in the global of education. Consequently, educational systems and curricula are under pressure to reform and equip students with these skills to ensure they are prepared for future challenges.

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 your analytical skills, but also nurtures your creativity and divergent thinking skills (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), collaborative and teamwork (Zheng et al., 2024). This approach aligns with the needs of the 21st-century learner by emphasizing personalized education paths that cater to individual strengths and preferences (Zhang et al., 2024). 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 endeavours. Additionally, incorporating CPS into the curriculum prepares students for the demands of the rapidly evolving workforce of the 21st century (Stankovic et al., 2017). As the world becomes increasingly complex and interconnected (Brunner et al., 2024), 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 can 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 can 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. Furthermore, research has explored 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 limited understanding of the specific factors that develop in CPS abilities of students.

Regarding assessment, previous studies have explored alternative approaches to assess CPS skills. Performance-based assessments, portfolios, and rubrics that assess creativity, critical thinking, metaphorical thinking, problem-solving related technology motivation, and problem solving skills have been proposed as more comprehensive and authentic assessment methods (Abosalem, 2016; Farida et al., 2022; Liu et al., 2024; Montgomery, 2002; Suastra et al., 2019). These approaches provide a more holistic view of the students' CPS skills and encourage the development of higher-order thinking abilities is limited. However, research also offers potential solutions, such as professional development of teachers, integration into 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 increasing recognition of the importance of CPS in independent curriculum education, there is a lack of comprehensive psychometric validation of 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 the CPS learning of students in the classroom. Therefore, this study aims to validate CPS using the Rasch model by investigating whether the data align with the measurement of the Rasch model. The reseach questions are followed:

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

1.1. CPS

CPS is a process that enables people to apply creative and critical thinking to find solutions to everyday problems (T. Lee et al., 2023; Van Hooijdonk et al., 2023). CPS helps

to eliminate the tendency to approach problems haphazardly and, as a result, prevent 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 such 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, organisation, leadership, and assessment, to discern the most significant problems for their respondents (Suryanto, Degeng, Djatmika & Kuswandi, 2021). Research states that creative problem solving can provide students with the 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 to 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.

CPS and ICT integration represent crucial intersections in contemporary education. In educational contexts, ICT serves as a powerful toolset that not only enhances traditional learning methods, but also fosters CPS skills among students (Guillén-Gámez et al., 2024; Mäkiö et al., 2022). Furthermore, ICT enables personalised learning experiences tailored to the needs of students (Gaeta, Miranda, Orciuoli, Paolozzi & Poce, 2013), empowering them to develop CPS skills in diverse and engaging ways (Andrews-Todd et al., 2023; Treffinger, 2007). As educational paradigms evolve, the integration of CPS and ICT not only prepares students for the challenges of the modern world but also equips them with essential skills to thrive in a digitally driven society.

Recent research underscores the importance of integrating ICT to improve CPS skills among students in educational settings. According to Selfa-Sastre, Pifarre, Cujba, Cutillas & Falguera (2022), ICT plays a crucial role in promoting creativity through collaborative problem solving and creative expression in language education. Their study highlights how digital technologies enable diverse learning opportunities and facilitate three key roles in enhancing collaborative creativity. These roles involve using interactive technologies to engage students in co-creative language learning experiences, equipping them with essential competencies to tackle complex challenges in a globalised and interconnected world. Moreover, Wheeler, Waite, & Bromfield (2002) emphasise that ICT enables students to engage in complex problem solving tasks that require creativity. Their findings suggest that integrating ICT tools into educational classrooms not only enhances students' technical skills but also cultivates their ability to think creatively and approach problems from different angles. These studies collectively underscore the synergy between CPS and ICT in education, highlighting ICT as a catalyst to nurture creative thinking and problem solving skills essential for 21st-century learners. Integrating ICT effectively into pedagogical practices not only enriches educational experiences but also prepares students to thrive in an increasingly complex and digital world.

Several researchers have developed instruments to assess CPS ability. For example, the study conducted by Hao et al. (2017) focused on developing a standardised assessment of CPS skills. Researchers recognised 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 importance 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 the CPS capabilities. They offer a comprehensive approach to measuring various aspects of CPS, including creative thinking, problem-solving strategies, and collaboration. Using these instruments, researchers and educators can gain insight into individuals' CPS abilities and tailor instructional strategies to enhance students' creative problem-solving skills.

1.2. Rasch measurement

Rasch measurement is a psychometric approach developed by Georg Rasch in the 1960s (Panayides et al., 2010). It is used to analyse and interpret data from educational and psychological assessments. The Rasch model, also known as the Rasch measurement model or the 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 (Tesio et al., 2023). In other words, probabilities are closely related to differences between item difficulty and individual ability (Boone et al., 2014). 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 the development of linear measures that are independent of the specific items used in the assessment (Caty et al., 2008). This means that the scores obtained from different sets of items can be compared and aggregated. The Rasch measurement also provides information about the reliability of the measurement and the fit of the data to the model, which helps to assess the quality of the assessment instrument. In educational and psychological research, the 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 evaluation, health outcomes research, and social sciences (Planinic et al., 2019). By employing the Rasch model, researchers can gain valuable insight into the relationship between individuals and items, refine measurement instruments, and make meaningful inferences about the construct being measured.

2. Methodology

2.1. Participants

In this cross-sectional study, a total of 137 higher education students participated as respondents. These students were selected from the Department of Mathematics Education in Indonesia using a stratified random sampling technique. The ethical approval is being considered by the Institutional Review Boards of the Universitas Islam Negeri Raden Intan Lampung. 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 other areas. The characteristics of the respondents are presented in Table 1.

Table 1

Characteristics of the 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
Living place		
City	69	50.4
Urban	68	49.6

2.2. Instrument

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 local curriculum in higher education to ensure that they effectively measure the desired skills and competencies. A total of 20 items were developed for this purpose, which encompass various aspects of CPS in the use of ICT.

The items in the instruments were carefully designed to assess the students' ability to apply higher-order thinking skills, critical and creative thinking, problem solving strategies, and collaboration within the context of real-world challenges. The items aimed to assess students' capacity to generate innovative solutions, think critically about complex problems, effectively communicate and collaborate with others, and demonstrate adaptability and resilience in problem-solving situations.

Using these instruments, the researchers aimed to obtain valuable insight into the CPS abilities of students and their ability to apply these skills in different scenarios. Instruments

were developed to provide reliable and valid measurement of CPS, enabling researchers to gain a comprehensive understanding of the strengths and areas for improvement in this domain. The use of these instruments in this study allowed for a systematic and standardised assessment of CPS, providing valuable data that can contribute to improving educational practices and curriculum development.

2.3. Procedure

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

2.4. Data analysis

The data collected in this study were analysed using Rasch measurement, a widely used psychometric approach to assess the fit between the observed data and the underlying measurement model. In this study, we used Winsteps version 4.7.0 (Linacre, 2020) to analyse the data, the Rasch model was applied to analyse 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. Data obtained from the CPS assessment instrument were analysed using various Rasch measurement parameters and techniques. The Outfit mean square (MNSQ) and Outfit z-standardised (ZSTD) were calculated to assess the fit of each item. The Outfit MNSQ provides an indication of the extent to which the observed responses deviate from the expected responses based on the Rasch model, while the Outfit ZSTD standardises the MNSQ values to facilitate comparison across items. The point-measure correlation (Pt-measure correlation) was computed to examine the relationship between the item difficulty and the person's ability. This correlation coefficient measures the strength of the association between item responses and estimated person abilities on the logit scale.

The Wright map, a graphical representation, was used 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 performed to identify the CPS abilities of the students. To explore possible differences in item functioning based on gender and living of residence, DIF analysis was performed. DIF analysis identifies items that may function differently for different groups, indicating potential bias or differential performance between groups. To analyse the differences in CPS abilities among students, SPSS version 26 was used. Descriptive statistics such as mean and SD were calculated to

provide an overview of the data. Additionally, R package statistics were employed to see the map.

3. Results

3.1. RQ1: Does the developed instrument demonstrate reliability and validity based on the Rasch measurement?

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

Table 2

The results 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%	

3.2. RQ2: What are the patterns of interaction between items and persons in the developed instrument based on the Wright map?

The pattern of interaction between items and individuals in the developed instrument, based on the Wright map, is presented in Figure 1. It can be seen in Figure 1 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 classified as a difficult item. The distribution or characteristics of difficult and easy items can be seen in Figure 2. On the other hand, the distribution of the fit of the items are shown in Figure 3.

Figure 1

Wright map

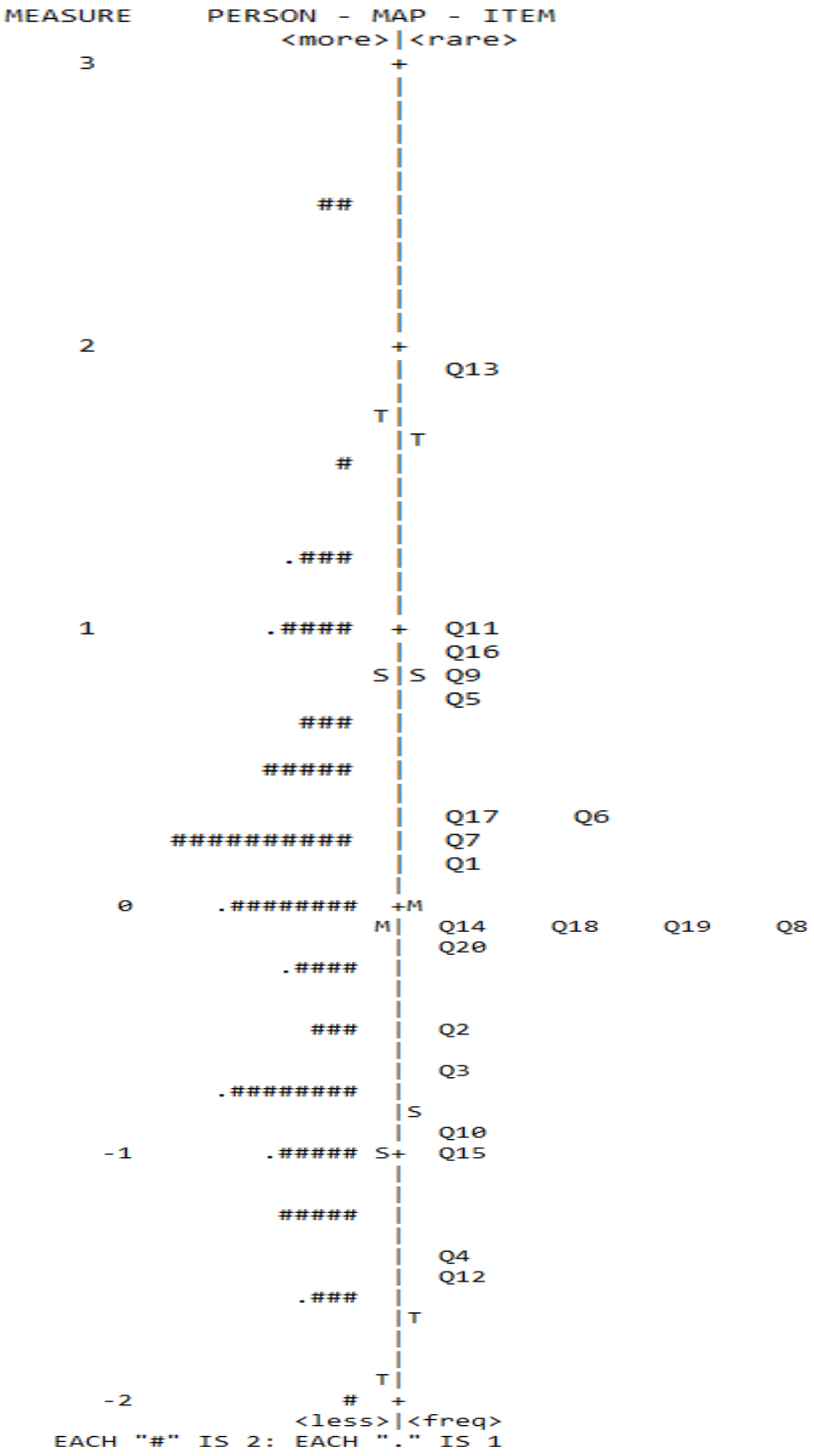


Figure 2

An item belongs to difficult and easier item

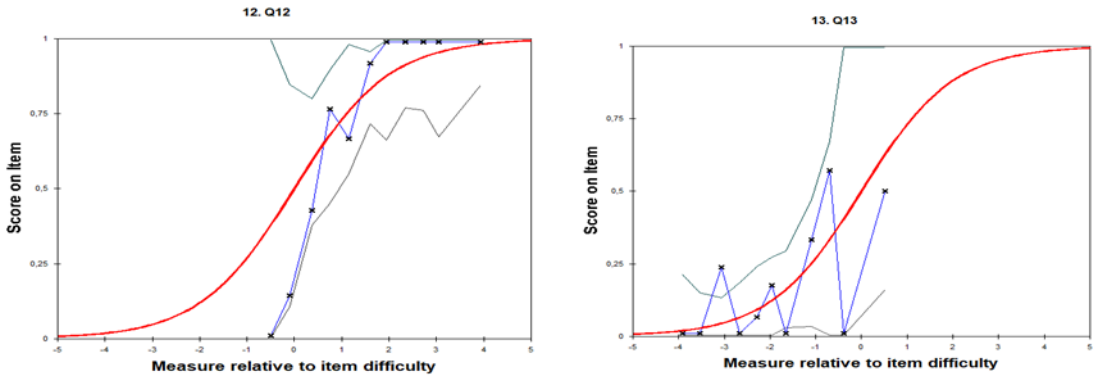
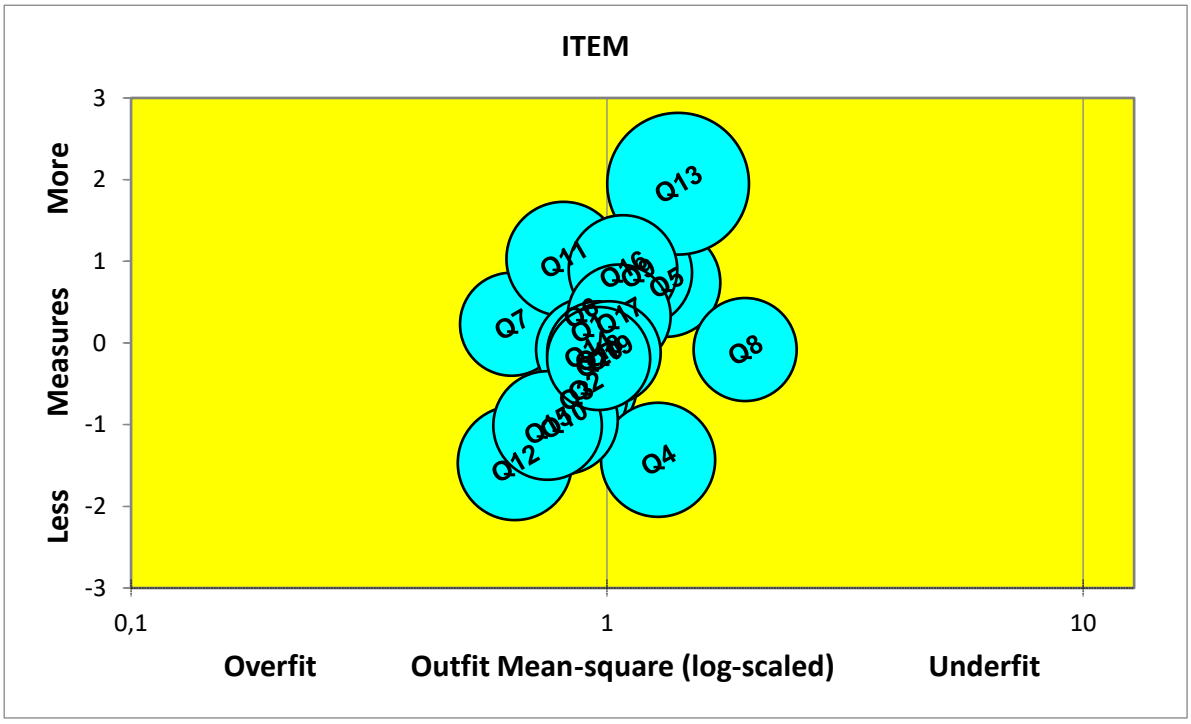


Figure 3

The distribution items are based on the Bubble Map



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 remove or revise item 8. As shown in Figure 5, Q8 appears to be approaching an underfit, indicating that it does not align well with the Rasch model.

3.3. RQ3: Are there any instrument biases based on gender according to the DIF analysis?

The DIF analysis was conducted to determine whether there were items that favoured 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-value is less than 0.05 or greater than -0.05 (Bond & Fox, 2015; Boone et al., 2014). Here are the results of the analysis using the Rasch model.

Table 4

Potential DIF owing gender

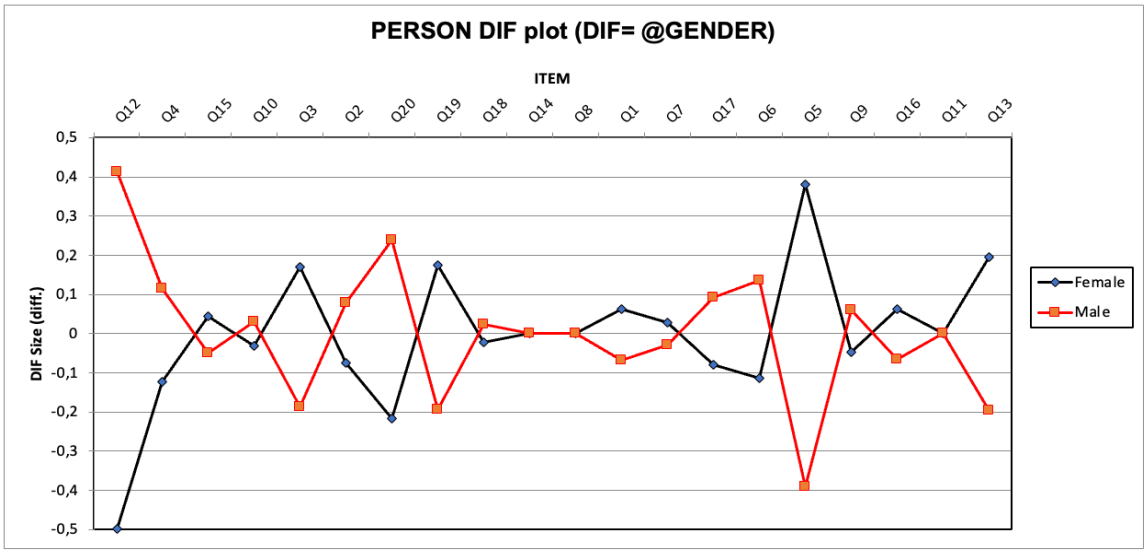
Item	DIF		DIF Contrast	t-value	Prob.
	Female	Male			
Q12	-1.97	-1.06	-0.91	-2.04	0.0432

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. 4) provides a visual representation of the DIF values for each item.

In the graph, the DIF values for item Q12 are noticeably higher compared to the 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 favour 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 revision of the item may be necessary to ensure a fair and unbiased assessment of all individuals, regardless of their gender.

Figure 4

Potential DIF owing gender



3.4. RQ 4: How does students' CPS develop in terms of course grades?

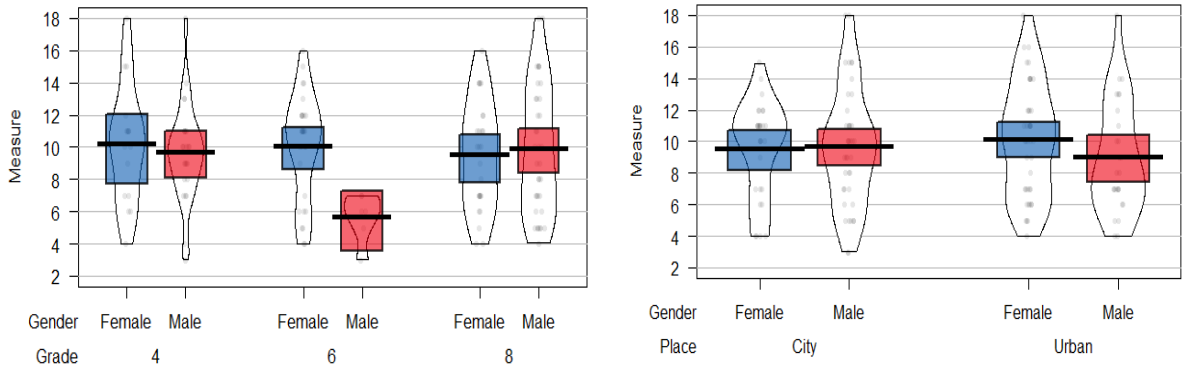
The statistical description of the students' responses to the given items is presented in Table 9. In Figure 5 (a), it can be observed that among female students, those in the fourth semester have an average response ability (M) of M = 10.19, SD = 3.73, followed by those

in the sixth semester with M = 10.00, SD = 3.32, and those in the eighth 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 versus rural), there are differences as shown in Figure 5 (b). 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 the collaborative problem-solving skills of students to some extent.

Figure 5

The students' ability to respond based on grade and gender (a), and grade and place of residence (b)



(a) (b)

4. Discussion

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 fits reasonably well with the Rasch measurement model, the ability to differentiate difficulty levels among items, and a sufficiently high level of reliability. The support of relevant research in this field also confirms these findings and provides a strong foundation to understand the measurement characteristics of this study. Previous studies, such as the one conducted by Soeharto (2021), also found similar results in terms of fit, separation, and reliability of the measurement.

The analysis results that indicate a 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 individual abilities measured. This is consistent with previous research stating that adequate separation is crucial to ensure reliable and valid measurement (Soeharto & Csapó, 2022).

Additionally, the reasonably good reliability for both items and individuals provides confidence that the measurement results obtained are reliable and consistent. In the context of this study, a reliability of 0.66 for items and 0.94 for persons indicates a satisfactory level of reliability. Other relevant studies can also provide support for the analysis of the 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. Furthermore, research by Avinç & Doğan (2024) found that the validity and relibality was confirmed by Rasch model. However, they noted that it would be advantageous to test its validity and reliability across various classes, age groups, and educational levels. Additionally, this study similar to Welter et al. (2024) that the psychometric properties has valid and reliabel using Rasch measurement. Overall, this analysis provides a deep understanding of the measurement characteristics of the items and individuals in this study. The support of 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 on the measured characteristics.

Based on the observed interaction pattern in Figure 3, 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. On the contrary, item 12 (Q12) is seen to be positioned higher, indicating that it belongs to the category of difficult items.

This analysis provides important information about the difficulty level of each element in the developed instrument. When the difficulty level of the items is known, adjustments and further development can be made to ensure that the items used cover appropriate levels of difficulty aligned with the research objectives. However, it should be noted that the assessment of 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 account other factors such as the characteristics of the respondents and the deeper context of the measurement. The items were effective in assessing the CPS abilities of students using ICT in the classroom (Wheeler et al., 2002) and impact on students digital competencies (Guillén-Gámez et al., 2024)This suggests that the difficulty level of the items was well-suited for the intended purpose of the instrument (Hobani & Alharbi, 2024). The information on item difficulty can guide further refinement and development of the instrument. It allows researchers to identify areas where the difficulty level may need to be adjusted, either by modifying existing items or adding new items to cover different difficulty levels.

Moving on to the Differential Item Functioning (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. 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 category of DIF. 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 a gender group.

However, it is important to note that the 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 help in instrument development and gain a better understanding of how the behavioural items in the instrument perform in specific groups. In other words, “DIF is not a synonym for bias,” as noted by Zieky (2012).

Items identified as DIF do not necessarily imply bias. According to Mollazehi & Abdel-Salam (2024), bias refers to the differing performance among individuals of equal ability from different subgroups due to irrelevant factors. DIF, introduced to distinguish the statistical meaning of bias from its social implications, focusses on the differing statistical properties that items exhibit among subgroups after matching individual abilities (Angoff, 2012). Since DIF interpretation is limited to differences in statistical properties, such as item difficulty and discrimination, expert panel reviews are necessary to determine if DIF items are biased (H. Lee & Geisinger, 2014).Thus, items showing DIF can be included in a test if no appropriate evidence of bias is found through panel review (De Ayala et al., 2002). In further research, steps can be taken to assess the causes of DIF and ensure that the instrument measures accurately without gender bias.

The statistical analysis presented in Figure 5 provides information on the ability of students to answer the given items based on the categories of gender, semester and residential location. The data gives an overview of the variation in the response abilities between different groups based on gender, semester, and residential location. Differences in means and standard deviations indicate variations in understanding of the material or learning approaches among these groups.

The performance of female students in the fourth semester, exhibiting an average ability score highest in creative problem-solving through the use of ICT, appears to surpass that of male students across different semesters. This observation aligns with previous research indicating that female students often demonstrate higher levels of proficiency in problem-solving tasks that require collaborative and ICT-related skills (Andrews-Todd et al., 2023; S. W.-Y. Lee et al., 2023). Studies have consistently shown that females tend to excel in collaborative learning environments, leveraging ICT tools effectively to enhance their problem-solving capabilities (Ma et al., 2023). This trend is attributed to various factors, including greater attention to detail, enhanced communication skills, and a preference for teamwork, which are critical in creative problem-solving tasks (Thornhill-Miller et al., 2023).

However, when comparing students living in urban and rural areas, differences can be observed in the results. Female students living in rural areas demonstrate the best performance in creative problem-solving through the use of ICT. Rural areas often face unique challenges such as limited access to resources, including educational infrastructure and technology (Alabdali et al., 2023). Despite these challenges, female students in rural areas may exhibit higher problem-solving abilities due to their adaptability and resilience in navigating these constraints. Research suggests that females often demonstrate higher levels of persistence and adaptability in learning environments (Dabas et al., 2023), which could contribute to their enhanced performance in creative problem-solving tasks involving ICT. Furthermore, cultural and societal factors may play a role in shaping educational outcomes (Min, 2023). In some cultures, there may be a stronger emphasis on education for females, particularly in rural settings where access to educational opportunities may be seen as transformative for individuals and families (Robinson-Pant, 2023). This emphasis could motivate female students to excel academically and in problem-solving tasks, including those involving ICT.

These findings highlight the potential influence of gender, semester, and residence location on the answering abilities of students. 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 response abilities.

Overall, these analyses provide valuable information on the characteristics of the measurement, the functioning of differential elements, 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.

5. Limitation and Future Research

This study provides important contributions to the development of evaluation instruments to evaluate 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 a better fit and validity of the evaluation instrument. Revision and refinement of items that do not meet the criteria is necessary to ensure that the evaluation instrument produces more accurate and reliable results.

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

The study also provides information on 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, no further explanation of the implications of these interaction patterns in the development of evaluation instruments is provided.

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

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

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.

6. Conclusions

On the basis of the analysis, the following conclusions can be drawn: Measurement using the Rasch model demonstrates a good fit between items and individuals in the evaluation instrument. The items exhibit a 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 more challenging. There are items that do not meet the DIF criteria based on gender. Item Q12 in the evaluation instrument tends to favour males over females in terms of the ability to answer. 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-centred instruction that aligns with the potential of students. Therefore, education can become more relevant, responsive, and enable learners to face the challenges of a complex world.

Authors' Contribution

Farida Farida: Conceptualization, Writing - Original Draft, Formal analysis; Yosep Aspat Alamsyah: Methodology, Editing, and Visualization; Bambang Sri Anggoro: Supervision, Funding acquisition, Writing – review & editing; Tri Andari: Formal analysis and Visualization; Restu Lusiana: Editing and Formal analysis.

Funding Agency

The Department of Research and Community Service (LP2M) funded the study reported at Universitas Islam Negeri Raden Intan Lampung, Indonesia (Grant number 99/2023).

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

Abina, A., Temeljotov Salaj, A., Cestnik, B., Karalič, A., Ogrinc, M., Kovačič Lukman, R., & Zidanšek, A. (2024). Challenging 21st-Century Competencies for STEM Students: Companies' Vision in Slovenia and Norway in the Light of Global Initiatives for Competencies Development. *Sustainability*, *16*(3), 1295. https://doi.org/10.3390/su16031295

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

Alabdali, S. A., Pileggi, S. F., & Cetindamar, D. (2023). Influential factors, enablers, and barriers to adopting smart technology in rural regions: A literature review. *Sustainability*, *15*(10), 7908. https://doi.org/10.3390/su15107908

Andrews-Todd, J., Jiang, Y., Steinberg, J., Pugh, S. L., & D'Mello, S. K. (2023). Investigating collaborative problem solving skills and outcomes across computer-based tasks. *Computers & Education*, *207*, 104928. https://doi.org/10.1016/j.compedu.2023.104928

Angoff, W. H. (2012). Perspectives on differential item functioning methodology. In *Differential item functioning* (pp. 3–23). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203357811-2/perspectives-differential-item-functioning-methodology-william-angoff

Arredondo-Trapero, F. G., Guerra-Leal, E. M., Kim, J., & Vázquez-Parra, J. C. (2024). Competitiveness, quality education and universities: The shift to the post-pandemic world. *Journal of Applied Research in Higher Education*. https://doi.org/10.1108/JARHE-08-2023-0376

Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the rasch model. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12662-7

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

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

Brunner, L. G., Peer, R. A. M., Zorn, C., Paulik, R., & Logan, T. M. (2024). Understanding cascading risks through real-world interdependent urban infrastructure. *Reliability Engineering & System Safety*, *241*, 109653. https://doi.org/10.1016/j.ress.2023.109653

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

Caena, F., & Redecker, C. (2019). Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (Digcompedu). *European Journal of Education*, *54*(3), 356–369. https://doi.org/10.1111/ejed.12345

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

Dabas, C. S., Muljana, P. S., & Luo, T. (2023). Female students in quantitative courses: An exploration of their motivational sources, learning strategies, learning behaviors, and course achievement. *Technology, Knowledge and Learning*, *28*(3), 1033–1061. https://doi.org/10.1007/s10758-021-09552-z

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

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

Gaeta, M., Miranda, S., Orciuoli, F., Paolozzi, S., & Poce, A. (2013). An Approach To Personalized e-Learning. *Journal of Education, Informatics & Cybernetics*, *11*(1). https://www.iiisci.org/Journal/pdv/sci/pdfs/HEB785ZU.pdf

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

Guillén-Gámez, F. D., Gómez-García, M., & Ruiz-Palmero, J. (2024). Digital competence in research work: Predictors that have an impact on it according to the type of university and gender of the Higher Education teacher:[Digital competence in research work: predictors that have an impact on

it according to the type of university and gender of the Higher Education teacher]. *Pixel-Bit. Revista de Medios y Educación*, *69*, 7–34. https://doi.org/10.12795/pixelbit.99992

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

Hobani, F., & Alharbi, M. (2024). A Psychometric Study of the Arabic Version of the "Searching for Hardships and Obstacles to Shots (SHOT)" Instrument for Use in Saudi Arabia. *Vaccines*, *12*(4), 391. https://doi.org/10.3390/vaccines12040391

Küçükaydın, M. A., Çite, H., & Ulum, H. (2024). Modelling the relationships between STEM learning attitude, computational thinking, and 21st century skills in primary school. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12492-7

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

Lee, D., & Lee, Y. (2024). Productive failure-based programming course to develop computational thinking and creative Problem-Solving skills in a Korean elementary school. *Informatics in Education*, *23*(2), 385–408. https://doi.org/10.1007/s10956-024-10130-y

Lee, H., & Geisinger, K. F. (2014). The Effect of Propensity Scores on DIF Analysis: Inference on the Potential Cause of DIF. *International Journal of Testing*, *14*(4), 313–338. https://doi.org/10.1080/15305058.2014.922567

Lee, S. W.-Y., Liang, J.-C., Hsu, C.-Y., Chien, F. P., & Tsai, M.-J. (2023). Exploring Potential Factors to Students' Computational Thinking. *Educational Technology & Society*, *26*(3), 176–189. https://doi.org/10.30191/ETS

Lee, T., O'Mahony, L., & Lebeck, P. (2023). Creative Problem-Solving. In T. Lee, L. O'Mahony, & P. Lebeck, *Creativity and Innovation* (pp. 117–147). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-8880-6_5

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

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

Liu, X., Gu, J., & Xu, J. (2024). The impact of the design thinking model on pre-service teachers' creativity self-efficacy, inventive problem-solving skills, and technology-related motivation. *International Journal of Technology and Design Education*, *34*(1), 167–190. https://doi.org/10.1007/s10798-023-09809-x

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

Ma, Y., Zhang, H., Ni, L., & Zhou, D. (2023). Identifying collaborative problem-solver profiles based on collaborative processing time, actions and skills on a computer-based task. *International Journal of Computer-Supported Collaborative Learning*, *18*(4), 465–488. https://doi.org/10.1007/s11412-023-09400-5

Mäkiö, E., Azmat, F., Ahmad, B., Harrison, R., & Colombo, A. W. (2022). T-CHAT educational framework for teaching cyber-physical system engineering. *European Journal of Engineering Education*, *47*(4), 606–635. https://doi.org/10.1080/03043797.2021.2008879

Min, M. (2023). School culture, self-efficacy, outcome expectation, and teacher agency toward reform with curricular autonomy in South Korea: A social cognitive approach. *Asia Pacific Journal of Education*, *43*(4), 951–967. https://doi.org/10.1080/02188791.2019.1626218

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

Mollazehi, M., & Abdel-Salam, A.-S. G. (2024). Understanding the alternative Mantel-Haenszel statistic: Factors affecting its robustness to detect non-uniform DIF. *Communications in Statistics - Theory and Methods*, 1–25. https://doi.org/10.1080/03610926.2024.2330668

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

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

Rahimi, R. A., & Oh, G. S. (2024). Rethinking the role of educators in the 21st century: Navigating globalization, technology, and pandemics. *Journal of Marketing Analytics*, 1–16. https://doi.org/10.1057/s41270-024-00303-4

Robinson-Pant, A. (2023). Education for rural development: Forty years on. *International Journal of Educational Development*, *96*, 102702. https://doi.org/10.1016/j.ijedudev.2022.102702

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

Selfa-Sastre, M., Pifarre, M., Cujba, A., Cutillas, L., & Falguera, E. (2022). The role of digital technologies to promote collaborative creativity in language education. *Frontiers in Psychology*, *13*, 828981. https://doi.org/10.3389/fpsyg.2022.828981

Soeharto, S. (2021). Development of a Diagnostic Assessment Test to Evaluate Science Misconceptions in Terms of School Grades: A Rasch Measurement Approach. *Journal of Turkish Science Education*, *18*(3), 351–370. https://doi.org/10.36681/tused.2021.78

Soeharto, S., & Csapó, B. (2022). Assessing Indonesian student inductive reasoning: Rasch analysis. *Thinking Skills and Creativity*, 101132. https://doi.org/10.1016/j.tsc.2022.101132

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

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., & Vidákovich, T. (2024). Relationship between ethnic identity, attitude, and mathematical creative thinking among secondary school students. *Thinking Skills and Creativity*, *51*, 101448. https://doi.org/10.1016/j.tsc.2023.101448

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

Tesio, L., Caronni, A., Kumbhare, D., & Scarano, S. (2023). Interpreting results from Rasch analysis 1. The "most likely" measures coming from the model. *Disability and Rehabilitation*, 1–13. https://doi.org/10.1080/09638288.2023.2169771

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., & Mourey, F. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence*, *11*(3), 54. https://doi.org/10.3390/jintelligence11030054

Treffinger, D. J. (2007). Creative Problem Solving (CPS): Powerful Tools for Managing Change and Developing Talent. *Gifted and Talented International*, *22*(2), 8–18. https://doi.org/10.1080/15332276.2007.11673491

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 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

Van Hooijdonk, M., Mainhard, T., Kroesbergen, E. H., & Van Tartwijk, J. (2023). Creative problem solving in primary school students. *Learning and Instruction*, *88*, 101823. https://doi.org/10.1016/j.learninstruc.2023.101823

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

Welter, V. D. E., Dawborn-Gundlach, M., Großmann, L., & Krell, M. (2024). Adapting a self-efficacy scale to the task of teaching scientific reasoning: Collecting evidence for its psychometric quality using Rasch measurement. *Frontiers in Psychology*, *15*, 1339615. https://doi.org/10.3389/fpsyg.2024.1339615

Wheeler, S., Waite, S. J., & Bromfield, C. (2002). Promoting creative thinking through the use of ICT. *Journal of Computer Assisted Learning*, *18*(3), 367–378. https://doi.org/10.1046/j.0266-4909.2002.00247.x

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

Yu, Y., & Duchin, F. (2024). Building a Curriculum to Foster Global Competence and Promote the Public Interest: Social Entrepreneurship and Digital Skills for American Community College Students. *Community College Journal of Research and Practice*, *48*(3), 164–174. https://doi.org/10.1080/10668926.2022.2064374

Zhang, L., Carter, R. A., Bloom, L., Kennett, D. W., Hoekstra, N. J., Goldman, S. R., & Rujimora, J. (2024). Are Pre-Service Educators Prepared to Implement Personalized Learning?: An Alignment Analysis of Educator Preparation Standards. *Journal of Teacher Education*, *75*(2), 219–235. https://doi.org/10.1177/00224871231201367

Zheng, J., Cheung, K., & Sit, P. (2024). The effects of perceptions toward Interpersonal relationships on collaborative problem-solving competence: Comparing four ethnic Chinese communities assessed in PISA 2015. *The Asia-Pacific Education Researcher*, *33*(2), 481–493. https://doi.org/10.1007/s40299-023-00744-y

Zieky, M. (2012). Practical questions in the use of DIF statistics in test development. In *Differential item functioning* (pp. 337–347). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203357811-20/practical-questions-use-dif-statistics-test-development-michael-zieky