



# European Journal of Educational Research

Volume 10, Issue 4, 2133 - 2144.

ISSN: 2165-8714

<http://www.eu-jer.com/>

## The Indonesian Version of the Physics Metacognition Inventory: Confirmatory Factor Analysis and Rasch Model

Moh. Irma

Sukarelawan\*<sup>ID</sup>

Universitas Negeri

Yogyakarta/Universitas

Ahmad Dahlan, INDONESIA

Jumadi<sup>ID</sup>

Universitas Negeri Yogyakarta,  
INDONESIA

Heru Kuswanto<sup>ID</sup>

Universitas Negeri Yogyakarta,  
INDONESIA

M. Anas Thohir<sup>ID</sup>

Universitas Negeri Malang,  
INDONESIA

Received: June 21, 2021 • Revised: September 22, 2021 • Accepted: October 2, 2021

**Abstract:** Metacognition inventory supports increased awareness and self-control to improve student's academic success, including physics. However, there are limitations to revealing the Physics Metacognition Inventory (PMI), especially in Indonesia. This study aims to explore and evaluate the psychometric properties of PMI. This survey research has involved 479 students from three high schools in Indonesia. The psychometric properties of the I-PMI were evaluated using a Confirmatory Factor Analysis and Rasch Model approach. The results show that the Indonesian Physics Metacognition Inventory (I-PMI) is collected in 6 constructs from 26 items. The validity, reliability, and compatibility tests have also been analyzed with good results. The five rating scales used have adequate functionality. This research has also presented more comprehensive information about the Physics Metacognition Inventory in the context of Indonesian culture. This study has implications for using I-PMI to assess students' metacognition at the high school level in Indonesia and recommendations for future research.

**Keywords:** *Confirmatory factor analysis, Physics Metacognition Inventory, psychometric evaluation, Rasch model, scale adaptation.*

**To cite this article:** Sukarelawan, M. I., Jumadi, Kuswanto, H., & Thohir, M. A. (2021). The Indonesian version of the physics metacognition inventory: Confirmatory factor analysis and Rasch model. *European Journal of Educational Research*, 10(4), 2133-2144. <https://doi.org/10.12973/eu-jer.10.4.2133>

### Introduction

Metacognition has become one of the central issues in educational research worldwide in the last 4 decades (Asy'ari et al., 2019; Kim et al., 2017; Zohar & Barzilai, 2013). The number of these studies has increased significantly since the 1970s (Harrison & Vallin, 2018). Positive influence on metacognition also occurs in almost all fields of science, including physics. Metacognitive awareness has significantly improved students' academic achievement (Hikmah et al., 2021; Taasoobshirazi & Farley, 2013; Uopasai et al., 2018). Students who have high metacognitive awareness tend to be able to improve learning outcomes (Panggayuh, 2017), conceptual understanding (Çetin, 2017; Taasoobshirazi et al., 2015), academic motivation (Mirzaei et al., 2012), and problem-solving abilities (Akben, 2020; Şahin & Kendir, 2013; Tachie, 2019). The next challenge is how to increase students' metacognitive awareness through effective problem-solving. This will be reflected in the learning vision through various models, including those in Indonesia.

Metacognition is an essential component that drives student success in problem-solving (Koyunlu Ünlü & Dökme, 2019). The term metacognition was first initiated by Flavell (1976) as an aspect of problem-solving. This is a higher-order thinking process that involves control and monitoring in learning (Livingston, 2003). Flavell conceptualizes metacognition as an individual's knowledge of cognitive activity or anything correlated with it (Coşkun, 2018; Cubukcu, 2009; Kim et al., 2017; Taasoobshirazi et al., 2015). In another view, metacognition is a person's activity or awareness in managing and controlling cognitive dynamics to have an effective way of learning (Sukarelawan & Sriyanto, 2019; Sulaiman et al., 2021). Then, metacognition develops in the context of problem-solving in collaborative groups (Graesser et al., 2018); skills and awareness (Akben, 2020); and success (Ali et al., 2018). However, there is limited availability of inventory to assess the impact of metacognition on students' problem-solving. With this awareness, Taasoobshirazi and Farley (2013) initiated a metacognitive inventory by developing a 24-item Physics Metacognition Inventory (PMI).

\* Corresponding author:

Moh. Irma Sukarelawan, Universitas Ahmad Dahlan, Indonesia. ✉ [moh.sukarelawan@mpfis.uad.ac.id](mailto:moh.sukarelawan@mpfis.uad.ac.id)



This inventory involves an essential construct to validate the instrument. Taasoobshirazi and Farley (2013) report that Exploratory Factor Analysis (EFA) has provided good evidence of construct validity and grouped 24-item into six metacognitive constituent factors, namely: knowledge of cognition, planning, monitoring, evaluation, debugging, and information management. Then in 2015, they revised and further developed the PMI to 26-item. Confirmatory Factor Analysis (CFA) and Rasch analysis have been used to explore the psychometric properties of PMI. As with their first research, they involved college physics students as respondents. The PMI uses a five-point Likert-type scale. Based on the CFA results, 26 items that make up the six factors in the PMI have a good model fit. Rasch analysis provides further evidence of the suitability of the PMI model to the Rasch model. This indicates that the 26 items developed contribute to the 6 factors that underlie the PMI. Therefore, the PMI becomes a valid and reliable inventory for measuring metacognition in problem-solving college physics students.

PMI cannot be directly applied to different cultures even though it has extraordinary psychometric properties. Therefore, a cross-cultural adaptation process is needed. Several researchers have adapted PMI to other cultures (Haeruddin et al., 2020; Koyunlu Ünlü & Dökme, 2019). Koyunlu Ünlü and Dökme adapts PMI into Turkish. A total of 8 experts have been involved in the adaptation process and 458 students for empirical validation. Experts have recommended that the Turkish version of PMI is compatible with Turkish culture. Based on the CFA analysis, the Turkish version of the PMI has a good model fit and meets the criteria for validity and reliability. Then Haeruddin et al. (2020) adapted PMI into the Indonesian version. A total of 3 linguists and 715 physics students were involved. Linguists recommend using PMI because it is by Indonesian culture. EFA results on 26 items indicate the formation of six factors. The CFA results support a good model fit for the Indonesian version of the PMI. Sukarelawan et al. (2021) conducted an inventory of heat and temperature materials in other cases. They showed six constructs which are Knowledge of cognition, Planning, Information management, Monitoring, Debugging, and Evaluation. In addition, the Indonesian version of the PMI has good validity and reliability. However, those were limited to explore the psychometric properties of the PMI. These two studies have not reported DIF (Differential Item Functioning) and the validity of the rating scale in PMI. DIF needs to be reported to determine whether the items that make up the inventory have a bias towards specific attributes of the respondents. Meanwhile, the validity of the rating scale needs to be reported to see if the 5-Point Likert scale used is functioning properly.

Rasch analysis needs to be used as an integrated approach to providing more comprehensive information on the psychometric properties of the inventories developed (Ning, 2018). Combining CFA and Rasch analysis provides a richer and deeper insight into psychometric properties in exploring student PMI in Indonesia (I-PMI). In addition to having good validity and reliability and compatibility with the model, the Indonesian version of the PMI should be free from gender bias. This study also provides rating scale results that do not confuse students. This is an advantage of Rasch analytics which offers this facility. Therefore, this study aims to bridge the gap in the literature through the adaptation of PMI into Indonesian for high school students. The findings of this study offer new and significant information in the Indonesian version of the PMI psychometric properties, which teachers or practitioners can use in assessing students' metacognition.

## Methodology

### *Research Design*

We explored I-PMI through pre-existing constructs (Haeruddin et al., 2020; Koyunlu Ünlü & Dökme, 2019; Taasoobshirazi et al., 2015). The results of the inventory obtained a metacognitive component consisting of 26 items. These are then grouped into six factors (constructs), namely: knowledge of Cognition (K, 6 items), planning (P, 5 items), free-body diagrams (F, 4 items), monitoring (M, 4 items), debugging (D, 3 items), and evaluation (E, 4 items). The I-PMI uses a 5-point Likert scale from "Never suits me" rated 1 to "Always suits me" rated 5. We consulted three English language experts. They concluded that the I-PMI was by Indonesian culture but needed minor revision.

### *Respondents*

The questionnaire was piloted on students in public secondary schools to produce a significant inventory. The CFA sample size requires respondents to be at least five times more than the observed variables (Hair et al., 2014; Kyriazos, 2018). The use of 146 respondents for the CFA analysis of 26 items in the I-PMI is considered to have met the specified sample size. Rasch analysis to assess the model's goodness is recommended between 50-250 respondents (Ling Lee et al., 2020). Therefore, this study (Rasch analysis) involved 333 students in fulfilling the assumption of data stability. We also consider outlier and misfit data. All respondents came from three public high schools in Indonesia, which were selected using stratified random sampling. The characteristics of the respondents are shown in Table 1.

Table 1. Characteristics of Respondents

	Frequency	Percentage	Mean <sub>age</sub>	SD <sub>age</sub>
CFA				
Male	40	27%	16.73	0.72
Female	106	73%	16.52	0.62
Total	146	100%	16.58	0.65
Rasch				
Male	79	23.7%	16.64	0.71
Female	254	76.3%	16.51	0.69
Total	333	100%	16.56	0.70

#### Data collection and ethics

The data collection process in this study has been in collaboration with local school teachers. We have discussed with the teacher what to convey to prospective respondents. Before collecting data, the teacher was asked to explain the purpose of this research. Student participation is voluntary. Students are given the right to withdraw their responses without being penalized. Confidentiality and anonymity have also been guaranteed (Lee et al., 2018; Taasobshirazi et al., 2015). We ask teachers to inform students that students' responses will not directly impact their learning outcomes.

#### Analyzing of Data

Several statistical applications have been used to support the data analysis process. Microsoft Excel and IBM SPSS Statistics 24 were used to tabulate the raw data. Confirmatory Factor Analysis (CFA) using Lisrel 8.50 software and Rasch probability model using Winsteps 4.6.1 software.

The structure of the I-PMI factor was analyzed using CFA because the original version has reported the formation of 6 factors on the metacognitive dimension. It is based on the CFA application paradigm, which states that the relationship between the observed variable and the latent variable is known (Byrne, 2013). Model parameters were evaluated employing a maximum likelihood method. Before conducting the overall fit test of the model, we checked the results of the factor loading estimation through the t-value of the item (accepted if  $> 1.96$ ) and the Standardized Loading Factor item (SLF, accepted if  $> 0.50$ ) (Wijanto, 2008). The overall fit of the model is evaluated from several fit indices such as absolute measures (such as  $\chi^2/df$ , RMSEA, and SRMR), incremental measures (such as TLI), and parsimony fit measures (such as PNFI). The overall fit of the model is met if  $\chi^2/df \leq 3$ ; RMSEA  $\leq 0.08$ ; SRMR  $\leq 0.08$ ; TLI  $\geq 0.95$ ; and PNFI  $> 0.60$  (Hair et al., 2014).

Rasch's analytical model makes it possible to process ordinal data into interval data from questionnaires (Setiawan et al., 2018). This model can reveal people's behavior by representing the measured items (Habibi et al., 2019). The reliability of the instrument in terms of the Cronbach coefficient, the person reliability coefficient, and the item reliability. Person data is used to see the suitability of respondents statistically, while item data is to know the appropriateness of using items in the instrument (Setiawan et al., 2018). The separation index on persons and items needs to be evaluated to see the function of the measurement instrument (Boone et al., 2014).

The fit of the measurement model was evaluated according to the MNSQ Infit and Outfit statistical values. The model fit interval was accepted in the range of  $0.5 < \text{MNSQ} < 1.5$  (Setiawan et al., 2018). Cronbach coefficient is used to evaluate the level of consistency of internal reliability. Bias on I-PMI items is detected through Differential Item Functioning (DIF). An item has a bias towards gender attributes if it has a probability value below 5% (0.05) (Bond & Fox, 2015; Sumintono & Widhiarso, 2014). The validity of the rating scale is carried out to verify the rating options used in the I-PMI. A good rating scale is if each respondent can distinguish the ratings provided. The validity of the rating scale is evaluated based on five criteria. First, the observed counts used have a unimodal distribution. Second, the number of frequencies for each rating is at least 10. Third, the Observed Average increases monotonically. Fourth, MNSQ Outfit  $< 2.0$ . Finally, step calibration increases monotonically (Papini et al., 2020; Sumintono & Widhiarso, 2014).

## Findings

#### Confirmatory Factor Analysis

This study aims to explore the psychometric properties of the Indonesian version of I-PMI, which will be used for high school students. This study uses CFA as a grouping of factors and confirms the factors of previous studies. Before carrying out the overall model fit test, the Standardized Loading Factor (SLF) value has been used to see an indication of an offending estimate. Table 2 summarizes the item loading of the 26 items in each I-PMI factor. These results were then compared to previous studies, namely version 1 (Taasobshirazi et al., 2015) and version 2 (Haeruddin et al., 2020).

Table 2. Comparison of Item Loadings on the between Recent Studies

Item	Item Loading			Composite Reliability
	Version 1	Version 2	Version 3	
Factor 1. Knowledge of Cognition				
K1	0.58	0.49	0.74	0.91
K2	0.63	0.62	0.82	
K3	0.74	0.44	0.80	
K4	0.62	0.54	0.75	
K5	0.73	0.60	0.79	
K6	0.77	0.61	0.80	
Factor 2. Planning				
P1	0.47	0.77	0.77	0.85
P2	0.31	0.68	0.76	
P3	0.46	0.61	0.57	
P4	0.35	0.72	0.80	
P5	0.52	0.54	0.75	
Factor 3. Free-Body Diagrams				
F1	0.90	0.55	0.83	0.92
F2	0.91	0.62	0.87	
F3	0.84	0.64	0.88	
F4	0.45	0.47	0.87	
Factor 4. Monitoring				
M1	0.91	0.84	0.81	0.88
M2	0.91	0.58	0.80	
M3	0.81	0.59	0.83	
M4	0.71	0.56	0.80	
Factor 5. Debugging				
D1	0.96	0.77	0.57	0.70
D2	1.00	0.66	0.69	
D3	0.13	0.61	0.72	
Factor 6. Evaluation				
E1	0.95	0.86	0.82	0.88
E2	0.91	0.78	0.89	
E3	0.61	0.71	0.85	
E4	0.58	0.62	0.66	

Note: Version 1 = Taasobshirazi et al. (2015), Version 2 = Haeruddin et al. (2020), Version 3 = this study.

The CFA results show that all items do not indicate an offending estimate. The factor load value of each item is statistically significant. The items in the I-PMI show a strong correlation to the formed factors. This has been supported by the loading value of each item exceeding the predetermined criteria. The E2 item ("I rechecked my work after completing the physics problem") had the highest correlation to the factor. Meanwhile, items P3 ("Before solving physics problems, I ignored the information I didn't need in the problem") and D1 ("After finishing physics questions, I double-checked my answers") had the lowest correlations among the other items. Composite reliability is a better approach in measuring internal consistency than Cronbach's alpha because it displays the standardized loadings of the manifest variables (Fornell & Larcker, as cited in Zulherman et al., 2021, p. 1702). The composite reliability of each factor/construct is between 0.70 – 0.92 ( $\geq 0.70$ ). The composite reliability value range shows that the internal consistency has been met. So, the I-PMI adapted has adequate validity and reliability (internal consistency).

The overall fit of the model was tested using the maximum likelihood method. Table 3 summarizes the overall fit index of the I-PMI model. Based on the overall fit index of the previously determined model, the values of  $\chi^2/df$ , RMSEA, SRMR, TLI, and PNFI have met the specified criteria. This indicates that the I-PMI model has a good fit.

Table 3. Comparison of the Model Fit Index with Previous Studies

Fit Parameter	Version 1	Version 2	Version 3
Sample Size	285	351	146
$\chi^2/df$	2.39	1.11	1.92
RMSEA	0.07	0.018	0.080
SRMR	0.08	-	0.053
TLI	-	-	0.97
PNFI	-	-	0.81
GFI	-	0.93	0.78
AGFI	-	0.92	0.72
CFI	0.93	0.99	0.98
NFI	-	0.93	0.96
IFI	0.93	-	0.98
P-value	< 0.05	> 0.05	< 0.05

Note: Version 1 = Taasobshirazi et al. (2015), Version 2 = Haeruddin et al. (2020), Version 3 = this study.

Rasch Analysis

Verification of the preferred rating in the I-PMI is carried out to test the validity of the rating scale. A good rating scale if the respondent does not feel confused about the ranking choices provided. Table 4 presents the results of the analysis of the I-PMI rating scale structure. The analysis results in Table 4 show that the number of frequencies on each rating scale is more than 10 and is unimodal. Observed average and step calibration Rasch-Andrich threshold increases monotonically. Outfit MNSQ scores from five categories are in the range of 0.80 - 1.47, below 2.0. In addition, the increase in step calibration > 1.0 scale. This is an evidence that the roles between response categories do not overlap. So overall, the I-PMI rating scale matches and functions well.

Table 4. Summary of I-PMI Rating Scale Function

Category	Count	Observed Average	Infit MNSQ	Outfit MNSQ	Step calibration Rasch-Andrich threshold
1	341	-0.74	1.40	1.47	NONE
2	1108	-0.45	0.93	0.94	-1.90
3	2962	0.26	0.83	0.80	-0.99
4	2733	1.20	0.82	0.86	0.80
5	1514	1.88	1.08	1.07	2.09

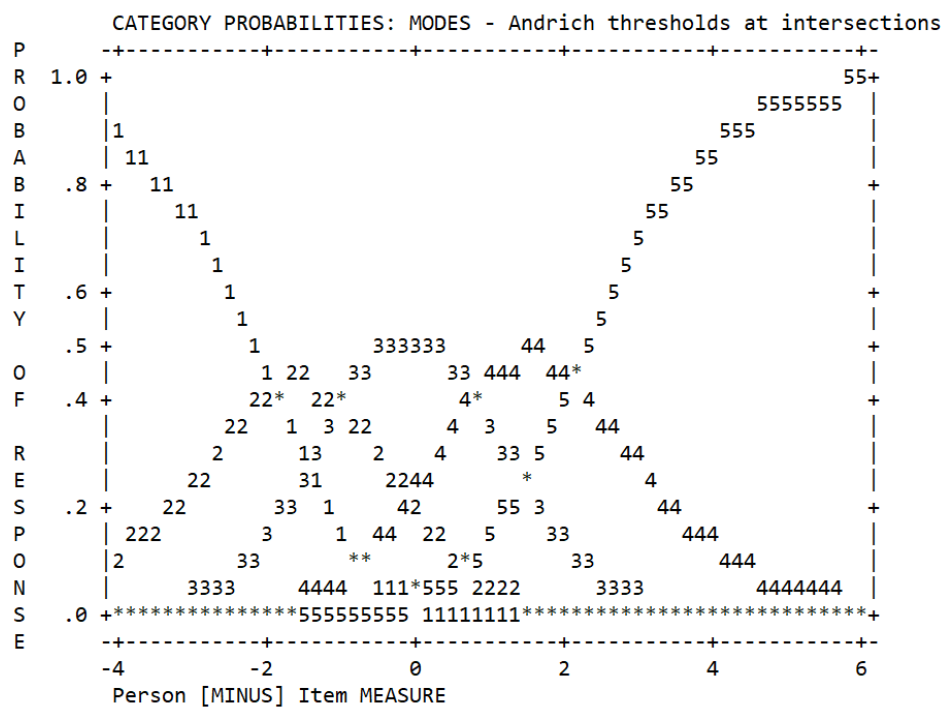


Figure 1. Category Probabilities curve for I-PMI

The summary of I-PMI statistics in Table 4 includes reliability and separation index on persons and items. Cronbach's  $\alpha$  value is obtained at 0.94 (excellent reliability). This shows the appropriate and reliable person-item interaction on the use of I-PMI. Person and item reliability values are 0.92 and 0.98, respectively. This value indicates the consistency of respondents in answering, and the quality of the items in the I-PMI is very good. The average person is 0.71 logit. This shows the tendency of respondents to agree on various attributes of the I-PMI. The person group (3.47) can be divided into 5 strata and the item group (7.06) into 10 strata based on the separation index. A high separation value indicates the ability of the inventory to identify groups of respondents and groups of items in various strata (Sumintono & Widhiarso, 2014).

Table 5. Summary of I-PMI Statistics

	Mean	SD	Separation	Reliability	Cronbach's $\alpha$
Person	0.71	1.06	3.47	0.92	0.94
Item	0.00	0.54	7.06	0.98	

Principal component analysis of residuals has been used to determine the unidimensionality of the I-PMI. This measure is essential to evaluate whether the I-PMI developed can measure metacognitive constructs. Scale unidimensionality is met if the raw variance explained by measures is not less than 40% and the unexplained variance does not exceed 15% (Sumintono & Widhiarso, 2014). Empirically, the raw variance explained by the measures is 45.7%, and the Unexplained variance is 8.3%. This value indicates that the I-PMI has good unidimensionality.

#### Item Fit

Table 6 summarizes the I-PMI's fit and outfit MNSQ statistics. Based on Table 6, it was found that item P3: "Before solving the physics problem, I ignored the information that I didn't need in the problem," which had an infit and outfit MNSQ statistic outside the acceptance value in the range of 0.5–1.5. A large MNSQ infit value indicates a less sensitive response pattern to the goal item on the person. Meanwhile, a large MNSQ Outfit value indicates a less sensitive response pattern to items with a certain level of difficulty in persons (Sumintono & Widhiarso, 2014). The MNSQ infit and outfit values on the P3 item indicate that the psychometric properties are inadequate in measuring the planning construct. This is in line with the results of the CFA analysis. Item P3 has an SLF value slightly above the acceptance limit value. So, item P3 needs to be improved. However, the overall MNSQ infit and outfit statistics have shown a good alignment between response patterns to goal items and item response patterns to certain difficulty levels (Pozo Muñoz & Bretones Nieto, 2019). This has supported the good validity of the I-PMI construct.

Table 6. Summary of Item Statistics

Item	Infit MNSQ	Outfit		PT-Mea Corr.	Measure (logit)	Model SE	Prob. (%)
		MNSQ	ZSTD				
K1	0.78	0.83	-2.42	0.62	0.34	0.07	13
K2	0.80	0.82	-2.52	0.63	0.46	0.07	22
K3	0.74	0.74	-3.75	0.66	0.45	0.07	3
K4	0.90	0.89	-1.55	0.66	0.31	0.07	4
K5	0.89	0.88	-1.61	0.66	0.31	0.07	6
K6	0.76	0.75	-3.61	0.70	0.23	0.07	8
P1	1.00	0.99	-0.07	0.65	-0.37	0.07	2
P2	0.93	0.94	-0.78	0.60	-0.47	0.07	47
P3	<b>1.93</b>	<b>1.99</b>	<b>9.90</b>	0.34	0.76	0.07	84
P4	0.92	0.90	-1.35	0.67	-0.30	0.07	1
P5	0.72	0.76	-3.45	0.70	-0.20	0.07	88
F1	1.12	1.12	1.55	0.63	0.88	0.07	10
F2	1.11	1.10	1.37	0.62	0.87	0.07	27
F3	1.13	1.12	1.62	0.65	0.72	0.07	3
F4	1.09	1.10	1.37	0.64	0.65	0.07	2
M1	1.04	1.00	0.06	0.65	-0.38	0.07	23
M2	0.77	0.77	-3.30	0.72	-0.19	0.07	8
M3	0.94	0.94	-0.84	0.67	-0.25	0.07	0
M4	1.03	1.03	0.40	0.67	-0.18	0.07	0
D1	<b>1.51</b>	1.47	<b>5.00</b>	0.45	-1.09	0.08	3
D2	<b>1.54</b>	1.49	<b>5.24</b>	0.45	-1.06	0.08	1
D3	0.78	0.79	-2.91	0.63	-0.27	0.07	100
E1	0.96	0.94	-0.82	0.66	-0.24	0.07	0
E2	0.75	0.75	-3.68	0.73	-0.19	0.07	2
E3	0.81	0.80	-2.82	0.71	-0.29	0.07	6
E4	1.05	1.01	0.21	0.65	-0.48	0.07	64

Differential Item Functioning

Table 6 shows an indication of the presence of DIF through the probability values of cross-gender items. A total of 12 out of 26 items (46%) were biased towards gender (numbers in bold). Nevertheless, the differences that are influenced by gender need to be examined more deeply. This can be done by paying attention to the tendency of DIF in the I-PMI, as shown in Figure 2. In the Knowledge of Cognition construct, items K3 and K4 are identified as having DIF for gender. They tend to be easy to agree with by the male group (see Figure 2a). Items P1 and P2 in the planning construct and items F3 and F4 in the free-body diagram construct have a DIF for gender. As in items K3 and K4, items P1, P2, F3, and F4 tend to be more difficult for women to agree with (see Figures 2b and 2c). In the monitoring construct, items M3 and M4 have DIF based on gender, and based on Figure 2d, it can be seen that there is an opposite trend with 6 items in the previous three constructs. Items M3 and M4 tend to be easier for female students to agree with than male. Two items (D1 and D2) in the Debugging construct and two items (E1 and E2) in the evaluation construct have DIF. The investigation results are in Figures 2e and 2f. These four items (D1, D2, E1, and E2) tend to be more difficult for male students to agree with. Therefore, biased items can be further selected to be discarded or retained.

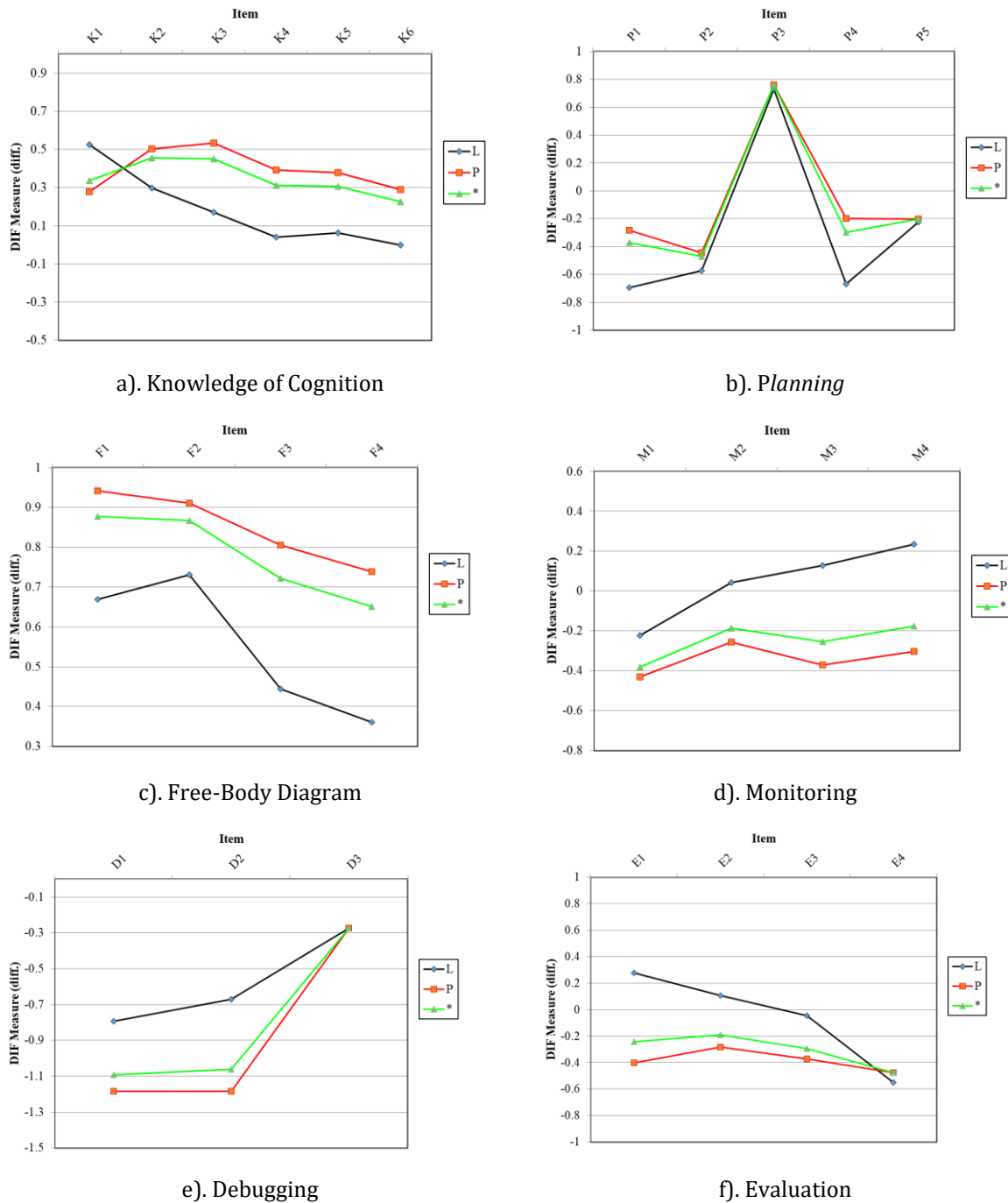


Figure 2. DIF Plot Item in I-PMI Based on Gender

Note: L = DIF measure of male, P = DIF measure of female, \* = average of DIF measure

Item and Person Distribution

Next, the person-item distribution is carried out to see the sensitivity and reliability of the I-PMI construct, as shown in Figure 3. The left section shows the distribution of persons based on their logit scores. Person logit scores ranged from -2.28 logit (for strongly disagree persons or students with low metacognitive levels) to +3.76 logit (strongly agree persons or students with high metacognitive levels). Students with high metacognitive levels are placed at the top of the map. The right side of the map shows the distribution of items by level of difficulty. The logit score ranges from -1.09 logit (item D1, with the lowest difficulty level or the easiest for students to agree on) to +0.88 logit (item F1, with the highest difficulty level or the most difficult for students to agree on). It appears that the average logit person is +0.71 higher than the average logit item (0.0 logit). This shows that students' metacognitive is above the average level of item standard difficulty. Several items need to be added to fill in the blanks in the interval -1.06 to -0.48 to provide optimal information when applied to low-ability students. This will have implications for increasing instrument sensitivity and reliability for respondents/students (Pozo Muñoz & Bretones Nieto, 2019).

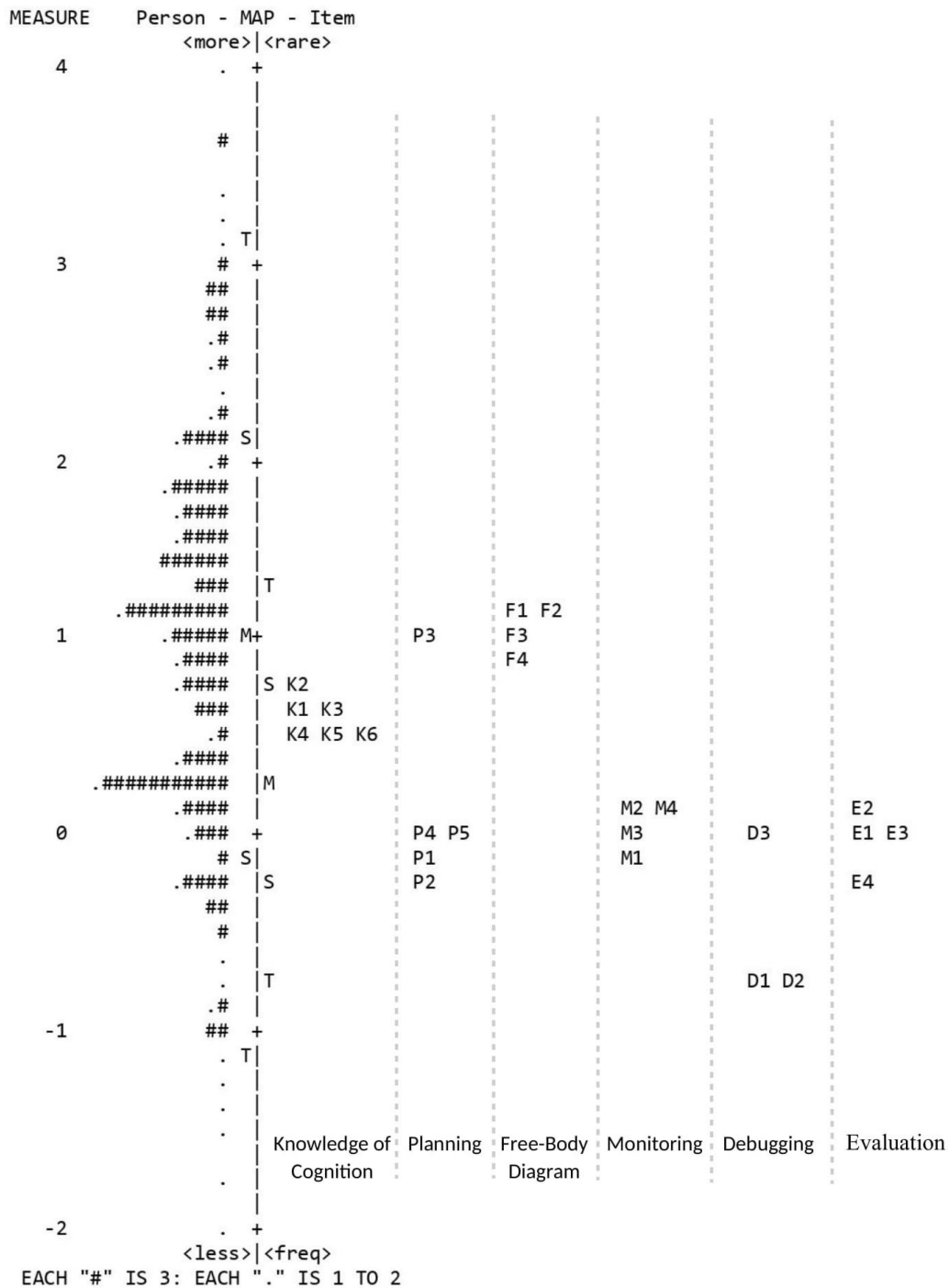


Figure 3. Wright Map of I-PMI



## Discussion

The focus of the research is exploring the physics metacognition inventory for high school students. Based on the analysis results, I-PMI has good psychometric behavior in high school students taking physics classes. The CFA and Rasch analysis results provide logical support for the 6 constructs/factors on the original scale. The results show that there are 26 items spread over 6 factors from the CFA analysis. This is also a suggestion that the two main components of metacognition, namely: Knowledge of cognition and Regulation of cognition (Schraw & Moshman, 1995; Taasoobshirazi et al., 2015; Zohar & Barzilai, 2013), can be broken down into six more specific ones. First, knowledge of cognition is knowledge, ideas, and theories that students have memorized about goals and strategies to complete tasks (Çetin, 2017; Zohar & Barzilai, 2013). It includes 3 sub-components: declarative, procedural, and conditional Knowledge (Dafik et al., 2019; Schraw & Moshman, 1995; Taasoobshirazi et al., 2015). Second, the regulation of cognition is a sequence of activities or actions students take to control their thinking or learning (Mahdavi, 2014). At the beginning of its development, regulation of cognition consisted of three components: planning, monitoring, and evaluation (Çetin, 2017; González et al., 2017; Kallio et al., 2017; Mahdavi, 2014). In its development, debugging and information management/free-body diagrams complement the regulation of cognition into five components (Asy'ari et al., 2019; Rahmat & Chanunan, 2018; Schraw & Dennison, 1994; Taasoobshirazi et al., 2015).

The results also show that the I-PMI in this study complements the previous results. These results align with the PMI structure reported by Taasoobshirazi et al. (2015) and Haeruddin et al. (2020). The comparison in Table 2 shows that the I-PMI has a factor structure by the model proposed by Taasoobshirazi et al. (2015) and supports the structure of factors formed in the research Haeruddin et al. (2020) in the Indonesian context. Therefore, this study strengthens the evidence for the suitability of using PMI across cultures and at different educational levels (Haeruddin et al., 2020; Koyunlu Ünlü & Dökme, 2019; Taasoobshirazi et al., 2015). Furthermore, the I-PMI is indicated to assess students' metacognition at the high school level. This research has presented a more comprehensive analysis by displaying CFA, Rasch model, Item Fit, DIF, and Pearson data distribution. This may serve as input for the construct developer to use all the tools so that the resulting model is in the students' metacognition measures inventory.

## Conclusion

The psychometric properties of the Indonesian version of the Physics Metacognitive Inventory (I-PMI) have been evaluated based on Confirmatory Factor Analysis and Rasch Model. The I-PMI adopted follows Indonesian culture and has good psychometric properties applied to state high school students in Indonesia. The CFA analysis has supported the 6-factor/construct structure on the original scale. The Rasch analysis shows the 5-point rating scale works well. However, some items in each construct have a gender bias.

## Recommendations

This study has supported the 6-factor structure of the metacognitive awareness scale in the Indonesian context. Due to these limitations, we recommend that future research considers the diversity of schools and across interests, cultural backgrounds, and even general education levels. A more comprehensive analysis can also be carried out by involving the equation model. This can clarify how the distinct components that weaken the validation can be detected. Further research can also further identify the factors in the I-PMI that contribute significantly to building students' awareness in managing their cognitive.

## Limitations

This research still has some limitations, including gender bias, not too wide age range, not involving private and cross-interest schools, and lack of heterogeneity in the cultural backgrounds of respondents. The results of the DIF show that the items that make up a scale are dependent on the gender in which the scale is developed (Williams et al., 2012). The emergence of bias threatens the validity of items in the I-PMI (Myers et al., 2006). I-PMI items are good if they are equal for each gender group (Papini et al., 2020). This may be because data collection was carried out in public schools only. The results cannot be generalized to students from private schools and across interests. However, this study has contributed to evaluating significant psychometric properties that were not reported by previous similar studies (Haeruddin et al., 2020; Koyunlu Ünlü & Dökme, 2019; Taasoobshirazi et al., 2015). These findings provide significant implications for practitioners, counselors, and parents to design and develop appropriate learning strategies for students.

## Acknowledgements

Thanks to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds in the doctoral dissertation scheme.

### Authorship Contribution Statement

Sekarelawan: Contribute to the concept and design aspects of research, data acquisition, data analysis, and drafting of manuscript. Jumadi: Contributed to the concept and design aspects, critical manuscript revision, supervision, and final approval. Heru Kuswanto: Contributed to aspects of critical manuscript revision, supervision, and final approval. Thohir: Strengthen discussion, admin, technical or material support.

### References

- Akben, N. (2020). Effects of the problem-posing approach on students' problem solving skills and metacognitive awareness in science education. *Research in Science Education*, 50(3), 1143–1165. <https://doi.org/10.1007/s11165-018-9726-7>
- Ali, M., Surif, J., Abdullah, A. H., Ibrahim, N. H., Talib, C. A., Shukor, N. A., Halim, N. D. A., Ali, D. F., & Suhairrom, N. (2018). The pattern of physics problem solving between more successful and less successful from metacognitive perspective. *Advanced Science Letters*, 24(11), 8476–8479. <https://doi.org/10.1166/asl.2018.12592>
- Asy'ari, M., Ikhsan, M., & Muhali. (2019). The effectiveness of inquiry learning model in improving prospective teachers' metacognition knowledge and metacognition awareness. *International Journal of Instruction*, 12(2), 455–470. <https://doi.org/10.29333/iji.2019.12229a>
- Bond, T., & Fox, C. M. (2015). *Applying the rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
- 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>
- Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge. <https://doi.org/10.4324/9781410600219>
- Çetin, B. (2017). Metacognition and self-regulated learning in predicting university students' academic achievement in Turkey. *Journal of Education and Training Studies*, 5(4), 132–138. <https://doi.org/10.11114/jets.v5i4.2233>
- Coşkun, Y. (2018). A study on metacognitive thinking skills of university students. *Journal of Education and Training Studies*, 6(3), 38–46. <https://doi.org/10.11114/jets.v6i3.2931>
- Cubukcu, F. (2009). Metacognition in the classroom. *Procedia - Social and Behavioral Sciences*, 1(1), 559–563. <https://doi.org/10.1016/j.sbspro.2009.01.101>
- Dafik, Suciarto, B., Irvan, M., & Rohim, M. A. (2019). The analysis of student metacognition skill in solving rainbow connection problem under the implementation of research-based learning model. *International Journal of Instruction*, 12(4), 593–610. <https://doi.org/10.29333/iji.2019.12438a>
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. R. Resnick (Ed.), *The Nature of Intelligence*. Lawrence Erlbaum Associates.
- González, A., Fernández, M.-V. C., & Paoloni, P.-V. (2017). Hope and anxiety in physics class: Exploring their motivational antecedents and influence on metacognition and performance. *Journal of Research in Science Teaching*, 54(5), 558–585. <https://doi.org/10.1002/tea.21377>
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, 19(2), 59–92. <https://doi.org/10.1177/1529100618808244>
- Habibi, H., Jumadi, J., & Mundilarto, M. (2019). The rasch-rating scale model to identify learning difficulties of physics students based on self-regulation skills. *International Journal of Evaluation and Research in Education*, 8(4), 659–665. <https://doi.org/10.11591/ijere.v8i4.20292>
- Haeruddin, Prasetyo, Z. K., Supahar, Sesa, E., & Lembah, G. (2020). Psychometric and structural evaluation of the physics metacognition inventory instrument. *European Journal of Educational Research*, 9(1), 215–225. <https://doi.org/10.12973/eu-jer.9.1.215>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson Education Limited.
- Harrison, G. M., & Vallin, L. M. (2018). Evaluating the metacognitive awareness inventory using empirical factor-structure evidence. *Metacognition and Learning*, 13(1), 15–38. <https://doi.org/10.1007/s11409-017-9176-z>
- Hikmah, F. N., Sukarelawan, M. I., Nurjannah, T., & Djumati, J. (2021). Elaboration of high school student's metacognition awareness on heat and temperature material: Wright map in Rasch model. *Indonesian Journal of Science and*

*Mathematics Education*, 4(2), 172–182. <https://doi.org/10.24042/ijsme.v4i2.9488>

- Kallio, H., Virta, K., Kallio, M., Virta, A., Hjärdemaal, F. R., & Sandven, J. (2017). The Utility of the metacognitive awareness inventory for teachers among in-service teachers. *Journal of Education and Learning*, 6(4), 78–91. <https://doi.org/10.5539/jel.v6n4p78>
- Kim, B., Zyromski, B., Mariani, M., Lee, S. M., & Carey, J. C. (2017). Establishing the factor structure of the 18-item version of the junior metacognitive awareness inventory. *Measurement and Evaluation in Counseling and Development*, 50(1–2), 48–57. <https://doi.org/10.1177/0748175616671366>
- Koyunlu Ünlü, Z., & Dökme, İ. (2019). Adaptation of physics metacognition inventory to Turkish. *International Journal of Assessment Tools in Education*, 6(1), 125–137. <https://doi.org/10.21449/ijate.483104>
- Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, 9(8), 2207–2230. <https://doi.org/10.4236/psych.2018.98126>
- Lee, Y. M., Song, J.-E., Park, C., & Son, Y.-J. (2018). Psychometric evaluation of the korean version of Patient-Centered care scale for hospital nurses. *Evaluation & the Health Professions*, 42(3), 344–365. <https://doi.org/10.1177/0163278718805244>
- Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. <https://doi.org/10.1177/2047487320902322>
- Livingston, J. A. (2003). *Metacognition: An overview*. ERIC. <https://files.eric.ed.gov/fulltext/ED474273.pdf>
- Mahdavi, M. (2014). An overview: Metacognition in education. *International Journal of Multidisciplinary and Current Research*, 2(6), 529–535.
- Mirzaei, F., Phang, F. A., Sulaiman, S., Kashefi, H., & Ismail, Z. (2012). Mastery goals, performance goals, students' beliefs and academic success: Metacognition as a mediator. *Procedia - Social and Behavioral Sciences*, 46, 3603–3608. <https://doi.org/10.1016/j.sbspro.2012.06.113>
- Myers, N. D., Wolfe, E. W., Feltz, D. L., & Penfield, R. D. (2006). Identifying differential item functioning of rating scale items with the rasch model: An introduction and an application. *Measurement in Physical Education and Exercise Science*, 10(4), 215–240. [https://doi.org/10.1207/s15327841mpee1004\\_1](https://doi.org/10.1207/s15327841mpee1004_1)
- Ning, H. K. (2018). A rasch analysis of the junior metacognitive awareness inventory with singapore students. *Measurement and Evaluation in Counseling and Development*, 51(2), 84–91. <https://doi.org/10.1080/07481756.2017.1358061>
- Panggayuh, V. (2017). Pengaruh kemampuan metakognitif terhadap prestasi akademik mahasiswa pada mata kuliah pemrograman dasar [The effect of metacognitive ability on student academic achievement in basic programming courses]. *Scientific Journal of Informatics Research and Learning/Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, 2(1), 20–25. <https://doi.org/10.29100/jipi.v2i1.228>
- Papini, N., Kang, M., Ryu, S., Griese, E., Wingert, T., & Herrmann, S. (2020). Rasch calibration of the 25-item Connor-Davidson resilience scale. *Journal of Health Psychology*, 26(11), 1976–1987. <https://doi.org/10.1177/1359105320904769>
- Pozo Muñoz, C., & Bretones Nieto, B. (2019). Spanish version of the flourishing scale (FS) on the parents of children with cancer: A validation through rasch analysis. *Frontiers in Psychology*, 10(35), 1–8. <https://doi.org/10.3389/fpsyg.2019.00035>
- Rahmat, I., & Chanunan, S. (2018). Open inquiry in facilitating metacognitive skills on high school biology learning: An inquiry on low and high academic ability. *International Journal of Instruction*, 11(4), 593–606. <https://doi.org/10.12973/iji.2018.11437a>
- Şahin, S. M., & Kendir, F. (2013). The effect of using metacognitive strategies for solving geometry problems on students' achievement and attitude. *Educational Research and Reviews*, 8(19), 1777–1792. <https://doi.org/10.5897/ERR2013.1578>
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475. <https://doi.org/10.1006/ceps.1994.1033>
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351–371. <https://doi.org/10.1007/BF02212307>
- Setiawan, B., Panduwangi, M., & Sumintono, B. (2018). A rasch analysis of the community's preference for different attributes of Islamic banks in Indonesia. *International Journal of Social Economics*, 45(12), 1647–1662. <https://doi.org/10.1108/IJSE-07-2017-0294>

- Sukarelawan, M. I., & Sriyanto, S. (2019). Mapping of profile students' metacognitive awareness in yogyakarta, Indonesia. *Journal of Research and Physics Education Research/ Jurnal Riset dan Kajian Pendidikan Fisika*, 6(2), 56–62. <https://doi.org/10.12928/jrkpf.v6i2.14556>
- Sukarelawan, M. I., Sulisworo, D., Jumadi, J., Kuswanto, H., & Rofiqah, S. A. (2021). Heat and temperature metacognition awareness inventory: A confirmatory factor analysis. *International Journal of Evaluation and Research in Education*, 10(2), 389–395. <https://doi.org/10.11591/ijere.v10i2.20917>
- Sulaiman, T., Syrene, S., Kai, W., & Subramaniam, P. (2021). Primary science teachers' perspectives about metacognition in science teaching. *European Journal of Educational Research*, 10(1), 75–84. <https://doi.org/10.12973/eu-jer.10.1.75>
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial* [Rasch model application for social sciences research]. Trim Komunikata Publishing House.
- Taasobshirazi, G., Bailey, M., & Farley, J. (2015). Physics metacognition inventory part II: Confirmatory factor analysis and rasch analysis. *International Journal of Science Education*, 37(17), 2769–2786. <https://doi.org/10.1080/09500693.2015.1104425>
- Taasobshirazi, G., & Farley, J. (2013). Construct validation of the physics metacognition inventory. *International Journal of Science Education*, 35(3), 447–459. <https://doi.org/10.1080/09500693.2012.750433>
- Tachie, S. A. (2019). Meta-cognitive skills and strategies application: How this helps learners in mathematics problem-solving. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(5), 1–12. <https://doi.org/10.29333/ejmste/105364>
- Uopasai, S., Bunterm, T., Muchimapura, S., & Tang, K. N. (2018). The effect of constructivism, metacognition and neurocognitive-based teaching model to enhance veterinary medicine students' learning outcomes. *Pertanika Journal of Social Sciences and Humanities*, 26(4), 2313–2331.
- Wijanto, S. H. (2008). *Struktural equation modelling dengan lisrel 8.8: Konsep dan tutorial* [Structural equation modeling with LISREL 8.8: Concepts and tutorials]. Graha Ilmu.
- Williams, B., Onsmann, A., & Brown, T. (2012). A rasch and factor analysis of a paramedic graduate attribute scale. *Evaluation and the Health Professions*, 35(2), 148–168. <https://doi.org/10.1177/0163278711407314>
- Zohar, A., & Barzilai, S. (2013). A review of research on metacognition in science education: Current and future directions. *Studies in Science Education*, 49(2), 121–169. <https://doi.org/10.1080/03057267.2013.847261>
- Zulherman, Zain, F. M., Napitupulu, D., Sailin, S. N., & Roza, L. (2021). Analyzing Indonesian students' google classroom acceptance during COVID-19 outbreak: Applying an extended unified theory of acceptance and use of technology model. *European Journal of Educational Research*, 10(4), 1697–1710. <https://doi.org/10.12973/eu-jer.7.3.555>