RASCH ANALYSIS TO EVALUATE THE PSYCHOMETRIC PROPERTIES OF JUNIOR METACOGNITIVE AWARENESS INVENTORY IN THE INDONESIAN CONTEXT

M. I. Sukarelawan*[1,6], Jumadi[2], H. Kuswanto[3], Soeharto[4], F. N. Hikmah[5]

[1,2,3]Postgraduate of Physics Education, Universitas Negeri Yogyakarta, Indonesia
[4]Doctoral School of Education, Faculty of Humanities and Social Sciences, University of Szeged, Hungary
[5]Department of Physics Education, Universitas Islam Negeri Antasari, Indonesia
[6]Postgraduate of Physics Education, Universitas Ahmad Dahlan, Indonesia

ABSTRACT

Empirically, metacognitive awareness is one of the main contributors to students' academic success. At the beginning of its development, the Jr.MAI self-report questionnaire was intended to measure students' metacognitive awareness in the United States. However, the evaluation of the psychometric properties for Indonesian high school students is still limited. The original Jr.MAI cannot be applied in Indonesia. By evaluating students' metacognitive awareness using Jr.MAI, teachers can understand students' information and knowledge related to their learning strategies and behaviors. Therefore, this study aims to evaluate the psychometric properties of the Indonesian translation of the Junior Metacognitive Awareness Inventory self-report questionnaire. The questionnaire consists of 18 items and uses a 5-point Likert scale response. 296 students (Male = 45.9%; Female = 54.1%) of public senior high schools in Indonesia completed the questionnaire. The Rasch model was used to evaluate the psychometric properties. The results showed that the 5-point rating scale with 18 items functioned properly with a good fit, no gender bias, and achieved unidimensionality and local independence assumptions. It proved that the Jr. MAI questionnaire defined the latent variables and classified persons and items properly. Therefore, we concluded that the developed questionnaire had good psychometric properties to be used by teachers and counselors for measuring and mapping the metacognitive characteristics at the senior high school level.

© 2021 Science Education Study Program FMIPA UNNES Semarang

INTRODUCTION

Various kinds of literature have disclosed the critical role of metacognition in students' learning process. The ability to monitor and control learning positively correlates with learning success, increased academic achievement, and students' health and well-being (Abdellah, 2015; Ning, 2018; Craig et al., 2020). It shows when metacognitive strategies positively correlate with students' test scores (Burin et al., 2020; Morphew, 2021) and when the practical use of metacognitive skills and strategies has improved learning at various levels of education (Abdellah, 2015; Herlanti, 2015; Ahdhianto et al., 2020; Amin et al., 2020).

Metacognitive taxonomy has evolved in the last four decades. Flavell is the initiator in introducing metacognitive concepts. At the beginning of its appearance, metacognitive was conceived as "thinking about cognitive phenomena" (Flavell, 1979). In other words, metacognitive is one's awareness or cognitive activity about the thinking process or everything related (Hidayat et

*Correspondence Address
E-mail: moh.irma2016@student.uny.ac.id

al., 2018). Flavell (1979) proposed a metacognitive structure consisting of four main components: metacognitive knowledge, metacognitive experiences, goal, and action. About a decade and a half later, Schraw and Dennison proposed a metacognitive structure with two main components: Metacognitive Knowledge (MK) and Metacognitive Experiences (ME). The MK component consists of three subcomponents: declarative, procedural, and conditional knowledge. At the same time, the ME component consists of five sub-components: planning, monitoring, information management, debugging, and evaluation (Schraw & Dennison, 1994). This proposal is a refinement of the factor structure proposed by Brown (1978).

In the Indonesian national education curriculum, students at the high school level must have metacognitive skills (Sukarelawan & Sriyanto, 2019). Therefore, a standard instrument is needed to facilitate the teachers in assessing these skills accurately. Several metacognitive questionnaires have been developed and applied in the field (Harrison & Vallin, 2018). This statement results from several previous metacognitive taxonomies (Flavell, 1979; Allen & Armour-Thomas, 1993; Schraw & Dennison, 1994). Because of their complex nature to observe and assess, self-report questionnaires are the most effective, efficient, and least complicated way to evaluate one's metacognitive awareness (Aydin & Ubuz, 2010; Craig et al., 2020). The component proposed by Schraw and Dennison has produced numerous self-report questionnaires to evaluate metacognitive content, for example, Metacognitive Awareness Inventory (MAI) (Schraw & Dennison, 1994), Junior Metacognitive Awareness Inventory (Jr.MAI) Versions A and B (Sperling et al., 2002), and Physics Metacognitive Inventory (PMI) (Taasoobshirazi & Farley, 2013; Taasoobshirazi et al., 2015).

Sperling et al. (2002) have developed Jr.MAI version B. The 18-item Jr.MAI has concurrent validity and forms two metacognitive factors (Knowledge of Cognition and Regulation of Cognition) based on exploratory factor analysis. The research report shows that the 18-item model fits the two factors (Table 3). The internal reliability of the Jr.MAI has a correlation coefficient of 0.82. Therefore, the items in Jr.MAI are reliable. These significant findings attracted the interest of researchers, including us, to reexamine the Jr.MAI in different contexts. Educational and psychological tests need a multilanguage version because interest in scientific achievement in international comparative studies and cross-cultural psychology has increased (Aydin & Ubuz, 2010). The Jr.MAI self-report instrument is intended to

measure students' metacognition in the United States, so the applications in different contexts need special attention. Several researchers have confirmed the use of Jr.MAI in other countries, for example, in Turkey (Aydin & Ubuz, 2010), Korea (Kim et al., 2017), and Singapore (Ning, 2018, 2019).

The use of metacognitive awareness instruments in Indonesia has been widely reported (Hidayat et al., 2018; Alindra et al., 2019; Fauzi & Sa'diyah, 2019; Sukarelawan & Sriyanto, 2019; Bahari et al., 2020; Yasir et al., 2020). However, literature reports on the adaptation process and comprehensive study of the instrument's psychometric properties, especially the Jr. MAI version B, are limited. Therefore, the study of the psychometric properties of Jr. MAI version B needs to be reported. This report will ensure the appropriateness and accuracy of the information in the Indonesian context.

Factor analysis techniques (exploratory factor analysis and confirmatory factor analysis) were used to establish the construct validity of Jr.MAI version B (Aydin & Ubuz, 2010; Kim et al., 2017; Ning, 2019). The rating scale function test, unidimensionality, bias analysis through Differential Item Functioning (DIF) test, and item quality (item difficulty and respondents' ability) are limitations not reported in previous studies. So, we need a Rasch analysis technique based on the item response theory to fill this gap. In our observations, supported by Craig et al. (2020), there are limited reports on the Rasch analysis to evaluate the use of psychometric properties of Jr.MAI in Indonesia. Therefore, this study aims to evaluate Jr. MAI's psychometric properties using the Rasch analysis technique in the Indonesian context. Thus, it is hoped that teachers or counselors can use Jr. MAI in measuring students' metacognition in Indonesia.

METHODS

The Indonesian version of Jr.MAI was administered to several senior high schools in Yogyakarta. The researchers chose Yogyakarta because it is the center of education known as a student city. Many students come from various regions in Indonesia to study in this city, so we assume the heterogeneity of students in Yogyakarta. Researchers asked permission from the school principal and teachers. Ethical approval was also granted from Yogyakarta State University. With the guidance and supervision from researchers and teachers, 351 students participated and filled out the online questionnaire using stratified

random sampling. Students were selected using the convenience sampling technique. We ran data screening to exclude outliers before data analysis. Fifty-five outliers were excluded from the dataset. Therefore, 296 students (Male = 45.9%; Female = 54.1%) were analyzed with Rasch measurement using Winsteps software. The number of participants in this analysis was more than 250, which was adequate for data stability for sample size (Chen et al., 2014; Lee et al., 2020).

The Jr.MAI instrument (Sperling et al., 2002) was adapted and translated into the Indonesian version using back-forward translation. The cross-cultural context was assessed in adapting the questionnaire in the Indonesian context (Muñiz et al., 2016). Jr.MAI questionnaire consists of 18 items divided into two constructs: Knowledge of Cognition (KoC) and Regulation of Cognition (RoC). In Jr.MAI, KoC and RoC consist of 9 items, respectively (Table 3). Each item uses five categories Likert scale starting from 1 (never) to 5 (always). Winsteps version 4.6.1 was employed to analyze the dataset based on Rasch measurement. We preferred to use Rasch measurement because it can solve some limitations of Classical Test Theory (CTT) analysis, such as missing data in the analysis, reliability parameter only using Cronbach's alpha, and the dependency of item and person, which may not be reliable and valid in another research context (Rusch et al., 2017).

Data analysis began with data screening of participant responses. We applied the rating scale analysis using Rasch modeling to perform data analysis. The rating scale function was reviewed from the increase in the average observation, Andrich Threshold values, and probability curves (Andrich, 2017; Van Zile-Tamsen, 2017). Instrument reliability was determined based on the Cronbach alpha coefficient, person and item reliability parameter, while item fit was determined from the Infit and Outfit MNSQ statistical value, the wright map, local independence, and unidimensionality. The bias of Jr.MAI items by gender is determined based on Differential Item Functioning (DIF) (Bond & Fox, 2015).

RESULTS AND DISCUSSION

The psychometric properties of the 18-item Jr.MAI were analyzed using WINSTEPS 4.6.1 software. The data analysis starts with a person screening and rating. In the data screening process, 55 outliers were detected. Outliers are students with suspicious and incompatible answers. Outfit MNSQ values were outside the acceptable criteria (0.5 to 1.5), indicating misfitting persons or outliers (Bond et al., 2015; Andrich, 2017). After person screening, we evaluated the rating scale used in Jr.MAI. Analysis of the choice function on the provided rating scale is critical as the element of the psychometric quality of the scale (Van Zile-Tamsen, 2017). A good rating scale is if the choices provided do not confuse the respondent. Table 1 shows the summary of the parameters used to assess the functioning of the options on the rating scale in Jr.MAI.

Table 1. Summary of Jr.MAI Rating Scale Function

Category	Counts	Observed Average	Andrich Threshold
1 (never)	37	-0.47	-
2 (Seldom)	265	-0.30	-2.77
3 (Sometimes)	1450	0.93	-1.38
4 (Often)	2439	1.90	0.94
5 (always)	1137	3.00	3.22

Based on Table 1, it appears that the observed counts have a unimodal distribution. The observed mean increased monotonically from -0.47 to +3.00 logit. Another indicator that needs to be considered to see the functioning of the scale choice is Andrich Threshold (Boone & Noltemeyer, 2017; Van Zile-Tamsen, 2017). The Andrich Threshold value increases monotonically from NONE to 3.22. A good choice of scale is if each level has increased by more than one logit in the Andrich Threshold parameter (Ning, 2018). There is an increase in each rating scale provided at least 1.39. Besides, the check of the scale function can be via a probability curve (Figure 1).
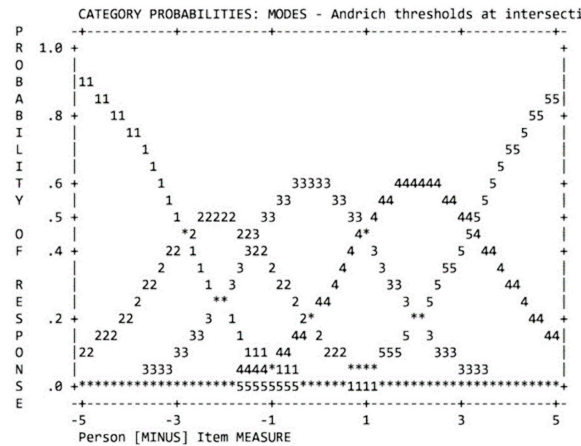
Figure 1. Probability Curves for the 5-point Likert Scale of Jr.MAI

All categories on the probability curve have their respective peaks along the Measure axis. It indicates a congruence with the recommended pattern (Rahayu et al., 2020). This finding is slightly different from the use of Jr.MAI in Singaporean students. The use of the Likert scale "Rarely" in Ning's (2018) study has a peak under the probability curve for the "sometimes" and "never" categories. For that reason, the Jr.MAI rating scale used for Singaporean students needs to be simplified to a 4-point Likert scale. Based on the average observation value, Andrich Threshold, and probability curves, it can be stated that the 5-point Likert scale used in Jr.MAI for Indonesian students can function properly.

After person screening and rating, we calculated the Jr.MAI questionnaire's summary statistics based on item and person parameters in Rasch modeling. Table 2 represents the Jr.MAI questionnaire's statistical summary for person and item based on Rasch parameters.

Table 2. Statistical Summary Based on Rasch Parameters

	Person	**Item**
N	296	18
Mean	68.7	1131
Measure	1.74	0
SD	0.87	0.92
SE	0.06	0.22
Mean Outfit ZSTD	-0.07	-0.10
Mean Outfit MNSQ	1.00	1.00
Separation	2.23	9.68
Strata	3.31	13.24
Reliability	0.83	0.99
Cronbach's Alpha	0.85	
Chi-squared ($\chi2$)	10072.06 (df= 10090)	
Probability	0. 5484 *	

*Normally distributed

Table 2 showed that the mean of person ability is 1.74, above the average level (0 logits), and item difficulties are in the average range (0 logits). Item separation indicated that the Jr.MAI questionnaire has 13 items in different difficulty levels. In this study, the person separation proved at least two person levels: the students with high and low ability. Overall, the data have Chi-squared ($\chi2$) = 10072.06 (df= 10090), p > 0.05 indicating normal distribution. Instrument reliability is estimated based on item and person (Table 2). The average person output is 1.74 logit, indicating a tendency for respondents to agree on various attributes in Jr.MAI. Data of item and person were used to see the suitability of the use of the item in Jr.MAI and the statistical suitability of

respondents. Person-item interaction on the use of Jr.MAI is appropriate and reliable because it has a Cronbach alpha value of 0.85. The respondents' consistency was good, and the quality of the items in Jr.MAI was exceptional (Didino et al., 2019). The result is supported by the value of the person and item reliability, respectively 0.83 and 0.99. These three reliability values indicate that the Jr.MAI items can define latent variables well (Maryati et al., 2019). The person and item separation value, represented in strata, is 3.31 and 13.24, respectively. This value indicates that Jr.MAI is proper to classify person and item.

The unidimensionality of the Jr.MAI scale is determined using the Principal Component Analysis of the residuals. The unidimensionality explains that the instrument is unidimensional in measuring latent factor in this study, Metacognitive Awareness. Jr.MAI can achieve validity criteria in measuring latent factor or unidimen-

sionality if the score of raw variance explained by measure is more than 30% (Chou & Wang, 2010). The value of raw variance explained by the Jr.MAI questionnaire measures is 42.8%. This value proves the existence of a good unidimensionality on the Jr.MAI scale measuring one dimension. These findings also support and reinforce the unidimensionality of Jr.MAI in Singapore students (Ning, 2018). The local independence explains that each Jr.MAI questionnaire item is not dependent. The instrument can achieve local independence criteria if the correlation between items is lower than 0.3. The raw residual correlation between items Jr.MAI questionnaire is below 0.3, proving that the questionnaire is free of local dependence issues.

The MNSQ infit and outfit statistical values in Table 3 were used to measure the suitability of individual items in Jr.MAI (Andrich & Marais, 2019).

Table 3. Comparison of Infit and Outfit MNSQ on Jr.MAI between Singapore Students (Ning, 2018) and Indonesian Students (This Study)

Item	Statement	Indonesian Students		Singapore Students		PT-Mea. Corr.	Measure
		Infit MNSQ	Outfit MNSQ	Infit MNSQ	Outfit MNSQ		
KoC1	I know when I understand something.	0.85	0.85	0.96	0.92	0.57	0.82
KoC2	When I like a topic, I study it more.	0.95	0.94	1.18	1.18	0.58	-0.75
KoC3	I pay attention to important information on a topic.	0.79	0.79	0.90	0.93	0.60	0.15
KoC4	I can make myself learn when I need to.	1.04	1.07	0.80	0.86	0.54	-0.11
KoC5	I can learn maximally when I already know something from the topic.	1.23	1.23	1.17	1.18	0.53	-0.41
KoC6	I understand what the teacher expects me to learn	1.28	1.30	1.41	**1.55**	0.43	2.22
KoC7	I will reuse the learning method that I have used successfully before.	1.12	1.12	0.85	0.88	0.44	-0.25
KoC8	I use my learning strengths to cover my weaknesses.	0.81	0.80	0.78	0.80	0.67	-0.53
KoC9	I sometimes immediately use learning strategies without a thinking process.	0.97	0.97	1.17	1.33	0.50	-0.31
RoC1	After finishing schoolwork, I ask myself if I have learned what I want to learn.	1.10	1.11	1.53	**1.64**	0.50	-0.10
RoC2	Before studying a topic, I think about what I want to learn from it.	0.94	0.95	0.97	1.05	0.57	-1.29
RoC3	When I learn new material, I ask myself how well I do it.	0.74	0.75	0.87	0.87	0.57	-0.19
RoC4	I consider several ways of doing my schoolwork and then choose the best way.	0.88	0.89	0.84	0.87	0.46	-0.11
RoC5	After finishing schoolwork, I ask myself if there is an easier way.	1.03	1.02	0.86	0.88	0.44	-0.79
RoC6	I make pictures or diagrams to help understand the material while studying.	0.82	0.83	0.86	0.87	0.49	0.80
RoC7	The learning strategies I use may differ depending on the works.	0.83	0.82	0.98	1.01	0.54	-1.10
RoC8	I make sure what to do before starting to do schoolwork.	1.24	1.22	1.14	1.14	0.54	-0.14
RoC9	I double-check my work to make sure it can be finished on time.	1.34	1.35	0.97	0.97	0.41	2.08

A fit item will make a good contribution in defining a common construct (Rahayu et al., 2020). In Rasch modeling, the ideal MNSQ infit and outfit value are 1. Value 0.5-1.5 is a reasonable acceptance range that shows the productive value for measurement (Wright & Linacre, 1994; Bond & Fox, 2015; Andrich & Marais, 2019). All 18 Jr.MAI items have infit values, and the MNSQ outfit is within the acceptance range. It shows a match between the response pattern to the target item and between person ability and item difficulty. This is different from the findings reported by Ning. Two items (KoC6 and RoC1) have insufficient psychometric properties (Table 3) (Ning, 2018). Besides, the PT-Measure Corr. value of the Jr.MAI scale moved in a positive direction from 0.41 to 0.67, as shown in Table 3. This shows the suitability of all items against the agreed latent variables (Maryati et al., 2019). Therefore, it can be concluded that the 18-item Jr.MAI can be app-

lied to measure the metacognitive awareness of high school students in Indonesia.

One of the characteristics of a good instrument is that it does not have a bias towards specific attributes of the respondents. DIF analysis was conducted to see the trend of Jr.MAI items on gender attributes. The items have a gender bias if the probability value is less than 5% (Sumintono & Widhiarso, 2014). As shown in Figure 2, there is no probability value less than 5% as an indication of item bias towards gender. We also ran a DIF analysis based on DIF size (Figure 3). Jr.MAI proved that there is no gender bias on each item because no DIF size has |DIF| score $\geq$ of 0.43 (slight to moderate) or |DIF| $\geq$ 0.64 logits (moderate to large) (Zwick et al., 1999). This result indicates that the Jr.MAI items are equal to the male and female, supporting other findings by Papini et al. (2020). These findings align with Jr.MAI in Singaporean students (Ning, 2018).
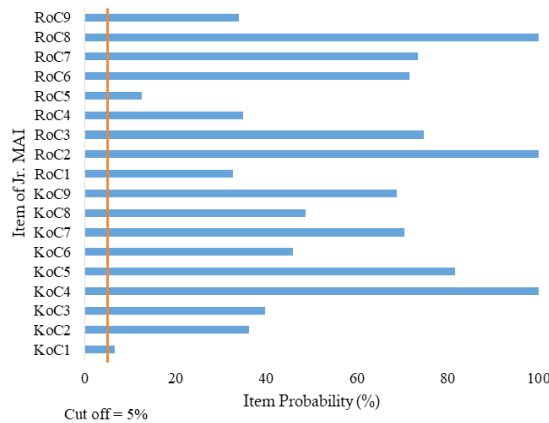
Figure 2. Item Probability Across Gender

Item KoC1 is near the cut-off criteria for probability value (p< 5%). However, we can assume that KoC1 is still worth retaining in the

Jr.MAI questionnaire. Overall, Figure 2 showed no bias issue in all items.
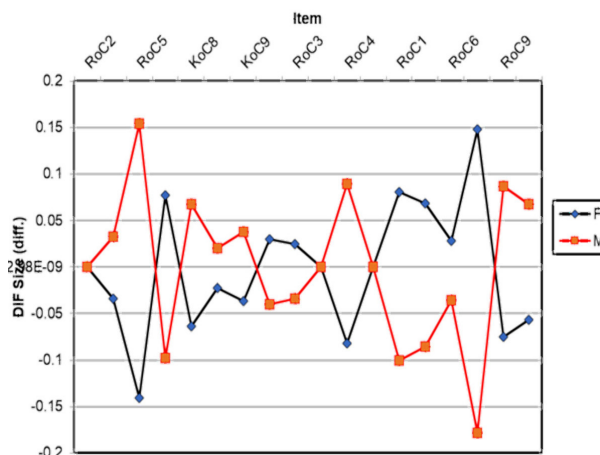
Figure 3. DIF Size Based on Gender (F=Female; M=Male)

We also performed a DIF analysis based on gender. Figure 3 attempted to illustrate the DIF in RoC Domain. However, Figure 3 indicated no substantial DIF size in the Jr.MAI questionnaire detected in item bias, less than 0.43 (slight to moderate categories).

The relationship between person and item is visualized through a person-item map (Wright map) (Wright & Stone, 1979). The Rasch model balances person ability and item difficulty on a linear scale in a one-frame reference, so the person ability and item difficulty are not interdependent (Wind & Gale, 2015; Blanc & Rojas, 2018). Figure 4 shows the state of the person and item on the same logit scale. It compares the item difficulty level against the person ability.

The person-item map in Figure 4 is divided into four areas. The upper-right area shows the position of the item with a high level of difficulty, or students tend to have more difficulty agreeing on the statement. Meanwhile, the lower-right area shows items with a low difficulty level, or students tend to agree with existing statements easily. The upper-left area shows the position of the person with a high metacognitive awareness level, and the lower-left area shows the position of the person with a low metacognitive awareness level. Eighteen items are distributed on the right side of the map. KoC6 item (I understand what the teacher expects me to learn) is the most difficult for students to agree on, and the RoC2 item (Before studying a topic, I think about what I want to learn from it) is the easiest for students to agree. There is a big gap between KoC3 and RoC9 items. Thus, some items need to be added to increase Jr. MAI's sensitivity and reliability (Muñoz & Nieto, 2019).

The mean score of the person is higher than the item. It shows that the students' average chance of having metacognitive awareness is higher than the average item difficulty level. Items and persons with the equivalent logit have a 50% probability of being agreed by the student. Items under a logit person have an agreed probability of more than 50%. At the same time, items with a logit above person have a probability of less than 50% to be agreed by students (Boone et al., 2014).
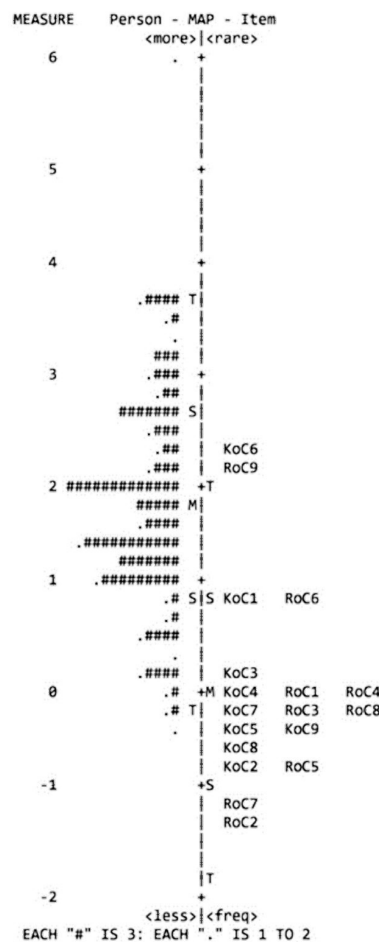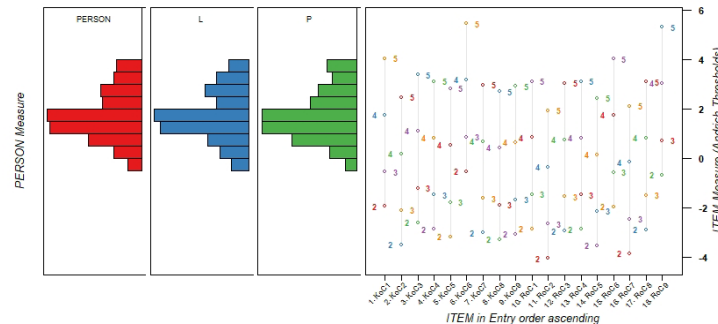

Figure 4. Person-item Map of the Jr.MAI

The person-item map indicated that students easily answer RoC7 and RoC2. However, those two items are still below two standard deviation logit units, so we can assume that these two items are not misbehaving items. This result indicated that students' regulation of cognition, especially RoC7 and RoC2, is higher than other item domains. There is room for improvement from the Jr.MAI person-item map by adding more difficult items to cover all person ability.

Figure 5. Person-item Map Based on Andrich Threshold

To ensure the Jr.MAI in the Indonesian context fits well based using rating scales, we ran a person-item map for the person and gender groups based on Andrich Threshold. Figure 5 gives us a further understanding of the distribution score in the questionnaire where all rating scales worked adequately for all person and gender in data fit and distribution.

CONCLUSION

Based on the analysis and discussion results previously presented, it was found that the use of the 5-point Likert scale in Jr.MAI is functioning well. The reliability of Jr.MAI is in a good category, and it can classify items and people from more than three groups. The 18 items fit well against the model and were free of gender bias. For that reason, Jr.MAI (Junior Metacognitive Awareness Inventory) has good psychometric properties to measure the metacognitive awareness of high school students in Indonesia. However, the item-person map indicated that there is still room for improvement to cover all students' ability by adding difficult items. No substantial bias was detected concerning response probability and DIF measures by gender. The limitation of this study is that it cannot be used on students who come from private schools because the respondents involved come from public schools. The students' cultural attributes were not included to see if Jr.MAI is free from bias towards culture. However, this study has made a significant contribution in evaluating the psychometric properties of Jr.MAI for use in Indonesia with the item response theory approach. The findings have significant implications for teachers, counselors, and parents to help students achieve academic success. Suggestions for future research are to focus on Differential Item Functioning. Therefore, future research needs to pay attention to the heterogeneity of the attributes of the respondents. For example, a researcher could assess Jr. MAI's bias towards school types (private and public schools), school location (urban and rural schools), or based on areas of interest (science and social fields).

ACKNOWLEDGEMENTS

REFERENCES

Abdellah, R. (2015). Metacognitive Awareness and its Relation to Academic Achievement and Teaching Performance of Pre-service Female Teachers in Ajman University in UAE. *Procedia - Social and Behavioral Sciences*, *174*, 560–567.

Ahdhianto, E., Marsigit, Haryanto, & Santi, N. N. (2020). The Effect of Metacognitive-Based Contextual Learning Model on Fifth-Grade Students' Problem-Solving and Mathematical Communication Skills. *European Journal of Educational Research*, *9*(2), 753–764.

Alindra, D., Fauzan, A., & Asmar, A. (2019). Metacognitive Awareness and Its Effect on Students' Problem Solving Ability in Implementing RME Approach. *Proceedings of the 1st International Conference on Innovation in Education*, *178*(ICoIE 2018), 318–322.

Allen, B. A., & Armour-Thomas, E. (1993). Construct Validation of Metacognition. *The Journal of Psychology*, *127*(2), 203–211.

Amin, A. M., Corebima, A. D., Zubaidah, S., & Mahanal, S. (2020). The Correlation between Metacognitive Skills and Critical Thinking Skills at the Implementation of Four Different Learning Strategies in Animal Physiology Lectures. *European Journal of Educational Research*, *9*(1), 143–163.

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

Andrich, D., & Marais, I. (2019). A course in Rasch measurement theory. *D. Andrich y I. Marais (Coords.), Measuring in the Educational, Social and Health Sciences*, 41-53.

Aydin, U., & Ubuz, B. (2010). Turkish Version of the Junior Metacognitive Awareness Inventory: An Exploratory and Confirmatory Factor Analysis Bilişüstü Yetiler Envanteri'nin Türkçeye Uyarlanması: Açımlayıcı ve Doğrulayıcı Faktör Analizi. *Education and Science*, *35*(157), 32–47.

Bahari, M., Widodo, A., & Winarno, N. (2020). Using Self-Regulated Learning to Promote Students 'Metacognitive Awareness in Learning Electricity.

Blanc, A., & Rojas, A. J. (2018). Use of Rasch person-item maps to validate a theoretical model for measuring Attitudes toward Sexual Behaviors. *PLoS one*, *13*(8), e0202551.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd ed.). Psychology Press.

Boone, W. J., & Noltemeyer, A. (2017). Rasch Analysis: A Primer for School Psychology Researchers and Practitioners. *Cogent Education*, *4*(1), 1416898.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer.

Brown, A. L. (1978). Knowing when, where, and how to remember; a problem of metacognition. *Advances in instructional psychology*, *1*.

Burin, D. I., Gonzalez, F. M., Barreyro, J. P., & Injoque-Ricle, I. (2020). Metacognitive Regulation Contributes to Digital Text Comprehension in E-learning. *Metacognition and Learning*, *15*(3), 391–410.

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch Model Analysis Applicable in Small Sample Size Pilot Studies for Assessing Item Characteristics? An Example using PROMIS Pain Behavior Item Bank Data. *Quality of Life Research*, *23*(2), 485–493.

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, *70*(5), 717–731.

Craig, K., Hale, D., Grainger, C., & Stewart, M. E. (2020). Evaluating Metacognitive Self-Reports: Systematic Reviews of the Value of Self-Report in Metacognitive Research. *Metacognition and Learning*, *15*(2), 155–213.

Didino, D., Taran, E. A., Barysheva, G. A., & Casati, F. (2019). Psychometric Evaluation of the Russian Version of the Flourishing Scale in a Sample of Older Adults Living in Siberia. *Health and Quality of Life Outcomes*, *17*(1), 1–12.

Fauzi, A., & Sa'diyah, W. (2019). The Metacognition of Pre-Service Biology Teachers: Awareness, Skills, Understanding, and Practices. *Proceedings of the 6th International Conference on Community Development (ICCD 2019)*, *January*, 26–32.

Flavell, J. H. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive–Developmental Inquiry. *American Psychologist*, *34*(10), 906–911.

Harrison, G. M., & Vallin, L. M. (2018). Evaluating the Metacognitive Awareness Inventory using Empirical Factor-Structure Evidence. *Metacognition and Learning*, *13*(1), 15–38.

Herlanti, Y. (2015). Kesadaran Metakognitif dan Pengetahuan Metakognitif Peserta Didik Sekolah Menengah Atas dalam Mempersiapkan Ketercapaian Standar Kelulusan pada Kurikulum 2013. *Cakrawala Pendidikan*, *34*(3), 357–367.

Hidayat, R., Zulnaidi, H., & Zamri, S. N. A. S. (2018). Roles of Metacognition and Achievement Goals in Mathematical Modeling Competency: A Structural Equation Modeling Analysis. *PLoS ONE*, *13*(11), e0206211.

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.

Lee, W. L., Chinna, K., & Sumintono, B. (2020). Psychometrics Assessment of HeartQol Questionnaire: A Rasch Analysis. *European Journal of Preventive Cardiology*, 2047487320902322.

Maryati, M., Prasetyo, Z. K., Wilujeng, I., & Sumintono, B. (2019). Measuring Teachers' Pedagogical Content Knowledge using Many-Facet Rasch Model. *Cakrawala Pendidikan*, *38*(3), 452–464.

Morphew, J. W. (2021). Changes in Metacognitive Monitoring Accuracy in an Introductory Physics Course. *Metacognition and Learning*, *16*(1), 89–111.

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Muñoz, C. P., & Nieto, B. 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.

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.

Ning, H. K. (2019). The Bifactor Model of the Junior Metacognitive Awareness Inventory (Jr. MAI). *Current Psychology*, *38*(2), 367–375.

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*, 135910532090476.

Rahayu, W., Putra, M. D. K., Iriyadi, D., Rahmawati, Y., & Koul, R. B. (2020). A Rasch and Factor Analysis of an Indonesian version of the Student Perception of Opportunity Competence Development (SPOCD) Questionnaire. *Cogent Education*, *7*(1), 1721633.

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.

Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, *19*(4), 460–475.

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of Children's Knowledge and Regulation of Cognition. *Contemporary Educational Psychology*, *27*(1), 51–79.

Sukarelawan, M. I., & Sriyanto, S. (2019). Mapping of Profile Students' Metacognitive Awareness in Yogyakarta, Indonesia. *Jurnal Riset Dan Kajian Pendidikan Fisika*, *6*(2), 56.

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Trim Komunikata.

Taasoobshirazi, 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.

Taasoobshirazi, G., & Farley, J. (2013). Construct Validation of the Physics Metacognition Inventory. *International Journal of Science Education*, *35*(3), 447–459.

Van Zile-Tamsen, C. (2017). Using Rasch Analysis to Inform Rating Scale Development. *Research in Higher Education*, *58*(8), 922–933.

Wind, S. A., & Gale, J. D. (2015). Rasch Measurement in the Context of a Misconceptions-Based Physical Science Assessment. *Science Education*, *99*(4), 721–741.

Wright, B. D., & Stone, M. H. (1979). Best test design.

Wright, B., & Linacre, J. M. (1994). Reasonable Mean-Square Fit Values. *Rasch Measurement Transactions*, *8*(3), 370.

Yasir, M., Fikriyah, A., Qomaria, N., & Al Haq, A. T. (2020). Metacognitive skill on students of science education study program: Evaluation from answering biological questions. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, *6*(1), 157–164.

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. *Journal of Educational Measurement*, *36*(1), 1–28.