MOH IRMA SUKARELAWAN moh.irma2016
<moh.irma2016@student.uny.ac.id>

Manuscript Update

1 pesan

Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>   10 Agustus 2021 15.42
Kepada: moh.irma2016@student.uny.ac.id

Dear Author,

We hope you are doing fine. We would like to inform you that there is a new regulation for articles that will be published in JPII, starting this year. Authors from Indonesia or domestic are asked to submit articles in the Indonesian version to be translated by parties who have been trusted by us so that the quality of the language in the articles to be published is even. Please submit the Indonesian version of the article on OJS using your account as the Author version and also to this email. If there is something unclear please reach us through this email. We are waiting for your response as soon as possible. We will wait for your Indonesian version until August 15, 2021. Thank you in advance.

Sincerely,
JPII Team

Jurnal Pendidikan IPA Indonesia

Principal Contact

Support Contact

Update: Manuscript Review

3 pesan

Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>   10 Agustus 2021 15.33
Kepada: moh.irma2016@student.uny.ac.id

Dear Authors,

We are pleased to inform you that the reviewer has uploaded the review results of your article.
Please check your OJS account for the newest review results.
We are looking forward to your revision **not later** than August 15, 2021.
Thank you.

Best regards,
JPII Team

Jurnal Pendidikan IPA Indonesia

- -

Principal Contact

Support Contact

MOH IRMA SUKARELAWAN moh.irma2016   14 Agustus 2021
<moh.irma2016@student.uny.ac.id>   09.36
Kepada: Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>

Yth. Editor,

Kami telah mengirimkan hasil revisi kedua dalam versi bahasa indonesia melalui author version.
Dengan senang hati kami menunggu proses review selanjutnya dalam rangka meningkatkan kualitas
paper kami.

Melalui email ini juga kami melampirkan hasil revisi terbaru dalam versi bahasa indonesia dan bahasa inggris.

Salam Hormat
Moh. Irma sukarelawan.

---

2 lampiran

📄 **04c. 27114-80900-2-RV Revisi 21_eng.docx**
699K

📄 **04c. 27114-80900-2-RV Revisi 21_indo.docx**
697K

Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>   14 Agustus 2021 10.02
Kepada: "MOH IRMA SUKARELAWAN moh.irma2016" <moh.irma2016@student.uny.ac.id>

Dear Author,

Thank you for your update.

Regards,
JPII Team

[Kutipan teks disembunyikan]

> [Kutipan teks disembunyikan]
> ---
> Untuk mendukung "Gerakan UNY Hijau", disarankan tidak mencetak email ini dan lampirannya.
> (To support the "Green UNY movement", it is recommended not to print the contents of this email and its attachments)
> # Universitas Negeri Yogyakarta
>   www.uny.ac.id
> ---

MOH IRMA SUKARELAWAN moh.irma2016
<moh.irma2016@student.uny.ac.id>

Update: Manuscript Review

1 pesan

Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>   14 April 2021 17.06
Kepada: moh.irma2016@student.uny.ac.id

Dear Authors,

We are pleased to inform you that the reviewer has uploaded the review results of your article.
Please check your OJS account for the newest review results.
We are looking forward to your revision **not later** than April 18, 2021.
Thank you.

Best regards,
JPII Team

Jurnal Pendidikan IPA Indonesia

- -

Nationally Accredited based on the Decree of the Minister of Research, Technology and Higher Education, Number 2/E/KPT/2015

Jurnal Pendidikan IPA Indonesia (Indonesian Journal of Science Education) [p-ISSN 2339-1286 | e-ISSN 2089-4392] published a scientific paper on the results of the study and review of the literature in the sphere of natural science education in primary education, secondary education and higher education. This journal in collaborate with *Perkumpulan Pendidik IPA Indonesia (PPII)* **/ Indonesian Society for Science Educators**

This journal has been indexed in Google Scholar, DOAJ, EBSCO, **SCOPUS**

Principal Contact

Parmin
Editor-in-Chief
Science Education Studies Program , Faculty of Mathematics and Natural Sciences, Semarang State University (UNNES)
D7 Building , 3rd Floor, Sekaran Campus, Gunungpati, Semarang, Indonesia 50229
Phone: 024-70805795
Fax: 024-8508005
Email: jpii@mail.unnes.ac.id

Support Contact

Parmin
Phone: +628164258038
Email: parmin@mail.unnes.ac.id

MOH IRMA SUKARELAWAN moh.irma2016
<moh.irma2016@student.uny.ac.id>

Update: Manuscript Review

4 pesan

Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>   26 Agustus 2021 09.38
Kepada: "MOH IRMA SUKARELAWAN moh.irma2016" <moh.irma2016@student.uny.ac.id>

Dear Authors,

We are pleased to inform you that the reviewer has uploaded the review results of your article.
Please check your OJS account for the newest review results.
We are looking forward to your revision **not later** than August 29, 2021.
Thank you.

Best regards,
JPII Team

Jurnal Pendidikan IPA Indonesia

**Nationally Accredited based on the Decree of the Minister of Research, Technology and Higher Education,
Number 2/E/KPT/2015**

Jurnal Pendidikan IPA Indonesia (Indonesian Journal of Science Education) [p-ISSN 2339-1286 | e-ISSN 2089-4392] published a
scientific paper on the results of the study and review of the literature in the sphere of natural science education in primary education,
secondary education and higher education. This journal in collaborate with *Perkumpulan Pendidik IPA Indonesia (PPII)* / **Indonesian
Society for Science Educators**

This journal has been indexed in Google Scholar, DOAJ, EBSCO, **SCOPUS**

Principal Contact

Parmin
Editor-in-Chief
Science Education Studies Program , Faculty of Mathematics and Natural Sciences, Semarang State University (UNNES)
D7 Building , 3rd Floor, Sekaran Campus, Gunungpati, Semarang, Indonesia 50229
Phone: 024-70805795
Fax: 024-8508005
Email: jpii@mail.unnes.ac.id

Support Contact

Parmin
Phone: +628164258038
Email: parmin@mail.unnes.ac.id

MOH IRMA SUKARELAWAN moh.irma2016   27 Agustus 2021
<moh.irma2016@student.uny.ac.id>   02.18
Kepada: Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>

Dear Editor.

We have made revisions to our manuscript based on suggestions from Reviewer C. All
suggestions from Reviewer C we transferred to our Indonesian version of the manuscript as

requested by the Editor at revision stage 2 (by email). We've also re-enabled the track changes feature so the Editor can track all the fixes that have been made.

We have uploaded the revised results via the author version in OJS

Best regards
Moh. Irma Sukarelawan
[Kutipan teks disembunyikan]

Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>   27 Agustus 2021 08.36
Kepada: "MOH IRMA SUKARELAWAN moh.irma2016" <moh.irma2016@student.uny.ac.id>

Dear Author,

Thank you for your update. You can use English only for the next revision. The article Review from Reviewer 3 has used the translated version. Please do the revision using that version. Thank you in advance.

Regards,
JPII Team

Jurnal Pendidikan IPA Indonesia

Principal Contact

Parmin
Editor-in-Chief
Science Education Studies Program , Faculty of Mathematics and Natural Sciences, Semarang State University (UNNES)
D7 Building , 3rd Floor, Sekaran Campus, Gunungpati, Semarang, Indonesia 50229
Phone: 024-70805795
Fax: 024-8508005
Email: jpii@mail.unnes.ac.id

Support Contact

Parmin
Phone: +628164258038
Email: parmin@mail.unnes.ac.id

[Kutipan teks disembunyikan]
    [Kutipan teks disembunyikan]

Universitas Negeri Yogyakarta
www.uny.ac.id

MOH IRMA SUKARELAWAN moh.irma2016   27 Agustus 2021
<moh.irma2016@student.uny.ac.id>   23.19
Kepada: Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>

Dear Editor,

Thanks for the information. We have resubmitted revisions based on suggestions from Reviewers C on the English version, and we have sent it in OJS. You can trace the revisions that we have made through the track changes feature in the paper. Thank you for this good communication.

Best regards
Moh. Irma Sukarelawan
[Kutipan teks disembunyikan]

MOH IRMA SUKARELAWAN moh.irma2016
<moh.irma2016@student.uny.ac.id>

Manuscript Update

2 pesan

Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>   10 November 2021 16.26
Kepada: "MOH IRMA SUKARELAWAN moh.irma2016" <moh.irma2016@student.uny.ac.id>

Dear Author,

Congratulations. Your article has been chosen to publish in JPII December 2021 Issue. Now that our publication payment is done before the final review process, so we would like to inform you about the publication fee. Please wait for our next instruction.

The publication fee, amounting to Rp 4.000.000,00 could be made to:

Bank Name: BNI
Bank Code: 009
Bank Address: Sekaran, Gunungpati, Kota Semarang
Account Number: 0031410331
Recipient Name: Universitas Negeri Semarang
Subject: JPII December 2021 Publication Fee
SWIFT No.: BNINIDJA

When you need to send in a currency other than Rupiah, please contact us first. We will wait for your payment until November 17, 2021. Please send the proof of publication payment by replying to this e-mail.

We are waiting for your response and congratulations once more.
Thank you for your cooperation.

All the best,

JPII Team

Jurnal Pendidikan IPA Indonesia

- -

Principal Contact

Phone: 024-70805795
Fax: 024-8508005
Email: jpii@mail.unnes.ac.id

Support Contact

Parmin
Phone: +628164258038
Email: parmin@mail.unnes.ac.id

MOH IRMA SUKARELAWAN moh.irma2016   14 November 2021
<moh.irma2016@student.uny.ac.id>   03.09
Kepada: Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>

Dear Editor in Chief of JPII

We are delighted to hear good news from you that our article has been accepted and published in the December 2021 issue.

For your information, we have followed up on the decision of the JPII editorial board regarding the acceptance of our article. We have made a payment of Rp. 4,000,000 to the BNI Semarang State University account. We have attached proof of payment to this email.

You can confirm back to me if there are things that need clarification (WA. 0895416066361).

Thank you.

Best regards
Moh. Irma Sukarelawan
[Kutipan teks disembunyikan]

JPII Fee.jpeg
151K

MOH IRMA SUKARELAWAN moh.irma2016
<moh.irma2016@student.uny.ac.id>

Letter of Acceptance and Publication Receipt

1 pesan

Jurnal Pendidikan IPA Indonesia <jpii@mail.unnes.ac.id>   24 Desember 2021 14.39
Kepada: "MOH IRMA SUKARELAWAN moh.irma2016" <moh.irma2016@student.uny.ac.id>

Dear Authors,

We are pleased to send the LoA and publication receipt of your manuscript.
It has been a pleasure to work with you.
Thank you very much.

Best regards,
JPII Team

Jurnal Pendidikan IPA Indonesia

Nationally Accredited based on the Decree of the Minister of Research, Technology and Higher Education, Number 2/E/KPT/2015

Jurnal Pendidikan IPA Indonesia (Indonesian Journal of Science Education) [p-ISSN 2339-1286 | e-ISSN 2089-4392] published a scientific paper on the results of the study and review of the literature in the sphere of natural science education in primary education, secondary education and higher education. This journal in collaborate with **Perkumpulan Pendidik IPA Indonesia (PPII) / Indonesian Society for Science Educators**

This journal has been indexed in Google Scholar, DOAJ, EBSCO, **SCOPUS**

Principal Contact

Parmin
Editor-in-Chief
Science Education Studies Program , Faculty of Mathematics and Natural Sciences, Semarang State University (UNNES)
D7 Building , 3rd Floor, Sekaran Campus, Gunungpati, Semarang, Indonesia 50229
Phone: 024-70805795
Fax: 024-8508005
Email: jpii@mail.unnes.ac.id

Support Contact

Parmin
Phone: +628164258038
Email: parmin@mail.unnes.ac.id

2 lampiran

M. I. Sukarelawan.jpeg
41K

M. I. Sukarelawan.jpg
841K

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

Moh. Irma Sukarelawan, Jumadi Jumadi, Heru Kuswanto,

Soeharto Soeharto, Fitri Nur Hikmah

Abstract

Early in its development, the Jr.MAI self-report questionnaire was intended for students in the United States. Recently, there has been increasing interest in developing multilingual tests in the educational and psychological fields. However, Studies were related to Jr.MAI in Indonesian context using Rasch measurement rarely conducted by researchers. Therefore, this study aims to evaluate the psychometric properties of the Indonesian translation of the Junior Metacognitive Awareness Inventory (Jr.MAI) self-report questionnaire. The Jr.MAI questionnaire is consisted of 18 items and used a 5-point Likert scale response. 296 students (Male = 45.9% and Female = 54.1%) of public senior high schools in Indonesia completed the Jr.MAI questionnaire. The Rasch model had been used to evaluate the psychometric properties of Jr.MAI. The results showed that the 5-point rating scale with 18 items was functioning properly with good fit, no gender bias and achieving the unidimensionality and local independence assumptions, which proved that Jr. MAI questionnaire defined the latent variables and classified people and items well. Therefore, we concluded that The Jr.MAI questionnaire developed had good psychometric properties to be used by teachers and counselors for measuring and mapping the metacognitive characteristics at the senior high school level.

Keywords: rasch analysis; Jr.MAI; metacognitive; rating scale model; Self-report questionnaire

INTRODUCTION

Today, various literatures have reported the important role of metacognition in the learning process of students. The ability to monitor and control learning has a positive correlation with learning success, increased academic achievement, and the health and well-being of students (Abdellah, 2015; Craig et al., 2020; Ning, 2018). For example, metacognitive strategies have a positive correlation with student test scores or student GPA (Vrugt & Oort, 2008; Young & Fry, 2008). Also, the empirical use of metacognitive skills and strategies has improved learning at various levels of education (Abdellah, 2015; Ahdhianto et al., 2020; Amin et al., 2020; Çetin, 2017; Herlanti, 2015).

Metacognitive taxonomy has evolved in the last 4 decades. Flavell has become an 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 can be viewed as awareness or cognitive activity of a person about the thought process or everything related to it (Hidayat et al., 2018). In 1979, Flavell (Flavell, 1979) proposed a metacognitive structure consisting of 4 main components (metacognitive knowledge, metacognitive experiences, goal, and action). About a decade and a half later, Schraw & Dennison proposed a metacognitive structure consisting of 2 main components, namely Metacognitive Knowledge (MK) and Metacognitive Experiences (ME). The MK component consists of three subcomponents, namely: declarative, procedural, and conditional knowledge. While the ME component consists of 5 sub-components, namely: planning, monitoring, information management, debugging, and evaluation (Schraw & Dennison, 1994). This proposal is a refinement of the factor structure proposed by Brown and Paris and colleagues (Craig et al., 2020).

In the Indonesian national education curriculum, students at the high school level are required to have metacognitive abilities (Sriyanto & Sukarelawan, 2019). Therefore, a standard instrument is needed that can facilitate the task of teachers in assessing these abilities accurately. Several metacognitive questionnaires have been developed and applied in the field (Harrison & Vallin, 2018). This is a result of the existence of several metacognitive taxonomies that have been proposed previously (Allen & Armour-Thomas, 1993; Flavell, 1979; Schraw & Dennison, 1994). Because of their difficult nature to observe and assess, self-report questionnaires are the most effective and efficient, and least problematic way to evaluate a person's metacognitive measures (Aydin & Ubuz, 2010; Craig et al., 2020).

The component proposed by Schraw & Dennison has produced numerous self-report questionnaires to evaluate metacognitive, 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 et al., 2015; Taasoobshirazi & Farley, 2013).

Sperling and colleague (Sperling et al., 2002) have developed Jr.MAI version B. Based on exploratory factor analysis, the 18-item Jr.MAI has concurrent validity and forms 2 metacognitive factors (Knowledge of Cognition and Regulation of Cognition). The research report shows that the 18-item model fits the two factors. 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). Given that the Jr.MAI self-report instrument is intended to measure metacognition of students in the United States, applications in different contexts need special attention. Several researchers have confirmed the use of Jr.MAI in different countries, for example in Turkey (Aydin & Ubuz, 2010), Korea (Kim et al., 2017), and Singapore (Ning, 2017, 2018).

Factor analysis techniques, exploratory factor analysis and confirmatory factor analysis, have been used to establish the construct validity of Jr.MAI (Aydin & Ubuz, 2010; Kim et al., 2017; Ning, 2017). Testing the rating scale function used, Jr.MAI unidimensionality, item bias analysis on respondent attributes through Differential Item Functioning (DIF) testing, item quality in constructing constructs associated with item difficulty level and respondent's ability are several limitations in using factor analysis in the literature (Aydin & Ubuz, 2010; Kim et al., 2017; Ning, 2017). So, we need a Rasch analysis technique based on the item response theory to fill this gap. As far as our observation is concerned, it has been supported by the systematic review results of Craig et al. from 1982 - 2018 (Craig et al., 2020), there have been no reports on the use of Rasch analysis to evaluate the psychometric properties of Jr.MAI use in Indonesia. Therefore, this study aims to evaluate the psychometric properties of Jr.MAI in Indonesian context using the Rasch analysis technique. Thus, it is hoped that Jr.MAI can be used by teachers or counselors in measuring and conducting metacognitive mapping of students in Indonesia.

METHODS

The initial participants in this study were 351 students in senior high school in Yogyakarta city, Indonesia. We ran data screening to exclude outliers before doing data analysis, 55 outliers excluded from the dataset. Therefore, 296 students (Male = 45.9% and Female = 54.1 were analyzed with Rasch measurement using Winsteps software. Data in this analysis were adequate in terms of data stability for sample size, more than 250 respondents (Jong et al., 2015; Ling Lee et al., 2020).

The original Jr.MAI instrument (Sperling et al., 2002) has been translated into Bahasa (Indonesian) by the Indonesian translator using back-forward translation. The cross-cultural context was assessed in adapting questionnaire in Indonesian context (Muñiz et al., 2016). Jr.MAI questionnaire consists of 18 items which are divided into 2 constructs, namely: knowledge of cognition (KoC) and regulation of cognition (RoC). In Jr.MAI, KoC and RoC each consist of 9 items. Each item uses a 5 category 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 using Rasch measurement because Rasch measurement can solve some limitations of Classical Test Theory (CTT) analysis such as missing data in the analysis, reliability parameter only using Cronbach's alpha, the dependency of item and person which may no be reliable and valid in another research context (Bradley et al., 2015)

Data analysis begins with an initial screening of the rating scale. The rating scale function is reviewed from the increase in the average observation, Andrich Threshold values, and probability curves (David Andrich, 2018; Sumintono & Widhiarso, 2014; Van Zile-Tamsen, 2017). Instrument reliability is determined based on the Cronbach alpha coefficient, person, and item reliability parameter. Item fit is determined from the Infit and Outfit MNSQ statistical value, the wright map, local independence, and unidimensionality. The bias of Jr.MAI items to gender is determined based on Differential Item Functioning (DIF) (Bond & Fox, 2015; Sumintono & Widhiarso, 2014).

RESULTS AND DISCUSSION

The psychometric properties of the 18-item Jr.MAI were analyzed using WINSTEPS 4.6.1 software. The data analysis begins with screening the person and rating scale. Table 1 represents the summary statistics of Jr.MAI questionnaire for person and item based on Rasch parameters.

Table 1. Statistical summary based on Rasch parameters

	Persons	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 ($\chi 2$)	10072.06 (df= 10090)	
Probability	0. 5484 *	

*Normally distributed

Screening Rating Scale
Analysis of the choice function on the provided rating scale is very important. A good rating scale is if the choices provided do not confuse the respondent. Table 2 shows a summary of the parameters used to assess the functioning of the options on the rating scale in Jr.MAI.

Table 2. 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 2, 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 (Sumintono & Widhiarso, 2014). The Andrich Threshold value increases monotonically from NONE to 3.22. A good choice of scale if each level has increased by more than 1 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). All Categories on the probability curve have their respective peaks along the Measure axis. This 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. So the Jr.MAI rating scale used for Singapore 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.
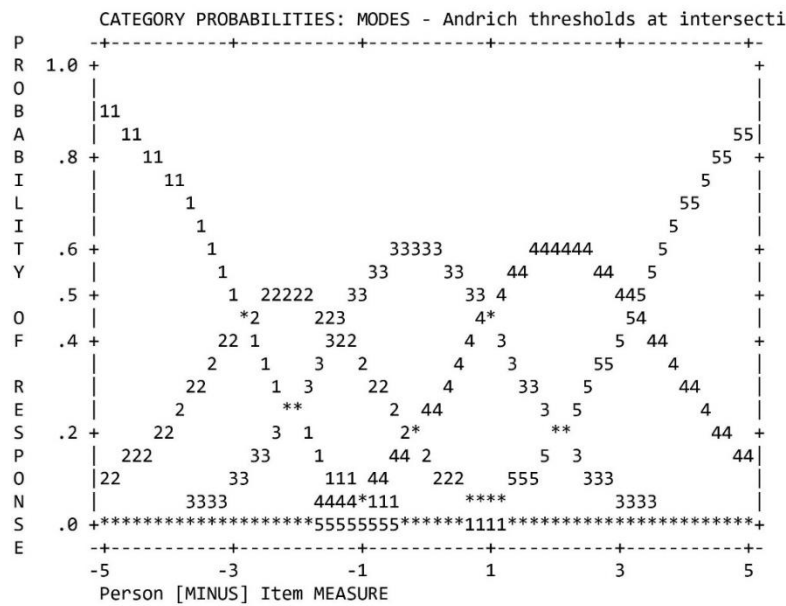
```
          CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersecti
     P    -+-----------+-----------+-----------+-----------+-----------+-
     R  1.0 +                                                           +
     O      |                                                           |
     B      |11                                                         |
     A      |  11                                              55|
     B   .8 +   11                                          55  |
     I      |     11                                         5   |
     L      |      1                                      55    |
     I      |       1                                    5      |
     T   .6 +        1          33333       444444     5         +
     Y      |         1        33     33    44    44  5          |
        .5 +          1  22222   33      33 4      445           +
     O      |           *2    223       4*       54            |
     F   .4 +          22 1     322      4  3      5  44         +
            |          2  1    3   2     4   3     55   4        |
     R      |         22    1 3    22    33   5    44           |
     E      |        2      **      2  44     3 5     4         |
     S   .2 +    22      3  1      2*        **        44     +
     P      |   222      33   1    44 2     5  3        44|
     O      |22         33       111 44    222   555   333       |
     N      |        3333      4444*111    ****       3333       |
     S   .0 +*******************55555555*****1111*******************+
     E      -+-----------+-----------+-----------+-----------+-----------+-
           -5          -3          -1           1           3           5
                    Person [MINUS] Item MEASURE
```

Figure 1. Probability curves for the 5-point Likert scale of Jr.MAI

Instrument Reliability

Instrument reliability is estimated based on items and persons (see Table 1). The average person output is 1.74 logit. This indicates a tendency for respondents to agree on various attributes in Jr.MAI. Data items and persons were used to see the suitability of using items 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 consistency of the answers from the respondents was good and the quality of the items in Jr.MAI was special (Cronbach, 1951; Didino et al., 2019). This is supported by the value of person and item reliability, respectively 0.83 and 0.99. These three reliability values indicate the items in Jr.MAI can define latent variables well (Maryati et al., 2019). The value of person and item separation, represented in the form of strata, is 3.31 and 13.24, respectively. This value indicates that Jr.MAI has a good ability to classify both person and item.

Unidimensionality and local independence

The unidimensionality of Jr.MAI scale has been determined using the Principal Component Analysis of the residuals. The unidimensionality explains that the instrument is unidimension in measuring latent factor in this study, Metacognitive Awareness. Jr.MAI can achieve validity criteria in measuring latent factor or unidimension if the score of raw variance explained by measure is more than 30% (Chou & Wang, 2010; Linacre, 1998). The value of raw variance explained by measures of Jr.MAI questionnaire 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 nature of Jr. MAI in Singapore students (Ning, 2018). The local independence explains that each item in the Jr.MAI questionnaire 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 which proves that the questionnaire is free of local dependence issues.

Item Fit

The MNSQ infit and outfit statistical values in Table 3 were used as measures of the suitability of individual items in Jr.MAI (D Andrich & Marais, 2019; Sumintono & Widhiarso, 2014). 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 (Andrich & Marais, 2019; Bond & Fox, 2015; Wright & Linacre, 1994). All 18 Jr.MAI items have infit values and the MNSQ outfit is within the acceptance range. This shows a match in the response pattern to the target item as well as a match between person ability and item difficulty (Sumintono & Widhiarso, 2014). Different from the findings reported by Ning. Two items (KoC6 and RoC1 have insufficient psychometric properties (see Table 2) (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 2. 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 applied to measure the metacognitive measures of high school students in Indonesia.

Table 3. Comparison of infit and outfit MNSQ on Jr.MAI between Singapore students (Ning, 2018) and this study.

Item	Indonesian students		Singapore students		PT-Measure Corr.	Measure
	Infit MNSQ	Outfit MNSQ	Infit MNSQ	Outfit MNSQ		
KoC1	0.85	0.85	0.96	0.92	0.57	0.82
KoC2	0.95	0.94	1.18	1.18	0.58	-0.75
KoC3	0.79	0.79	0.90	0.93	0.60	0.15
KoC4	1.04	1.07	0.80	0.86	0.54	-0.11
KoC5	1.23	1.23	1.17	1.18	0.53	-0.41
KoC6	1.28	1.30	1.41	1.55	0.43	2.22
KoC7	1.12	1.12	0.85	0.88	0.44	-0.25

KoC8	0.81	0.80	0.78	0.80	0.67	-0.53
KoC9	0.97	0.97	1.17	1.33	0.50	-0.31
RoC1	1.10	1.11	1.53	1.64	0.50	-0.10
RoC2	0.94	0.95	0.97	1.05	0.57	-1.29
RoC3	0.74	0.75	0.87	0.87	0.57	-0.19
RoC4	0.88	0.89	0.84	0.87	0.46	-0.11
RoC5	1.03	1.02	0.86	0.88	0.44	-0.79
RoC6	0.82	0.83	0.86	0.87	0.49	0.80
RoC7	0.83	0.82	0.98	1.01	0.54	-1.10
RoC8	1.24	1.22	1.14	1.14	0.54	-0.14
RoC9	1.34	1.35	0.97	0.97	0.41	2.08

Differential Item Functioning (DIF)

One of the characteristics of a good instrument is if it does not have a bias towards certain attributes of the respondent. DIF analysis was conducted to see the trend of items in Jr.MAI on gender attributes. 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 < 5% as an indication of grain bias towards gender. We also ran DIF analysis based on DIF size (see Figure 3). Jr.MAI proved that there is no gender bias on each item because no DIF size has |DIF| score ≥ 0.43 (slight to moderate) or |DIF| ≥ 0.64 logits (moderate to large) (Zwick et al., 1999). This result indicates the items in Jr. MAI are equal towards the male and female sex, which supports other findings from Papini et al. (2020). These findings are also in line with Jr. MAI in Singapore students (Ning, 2018).
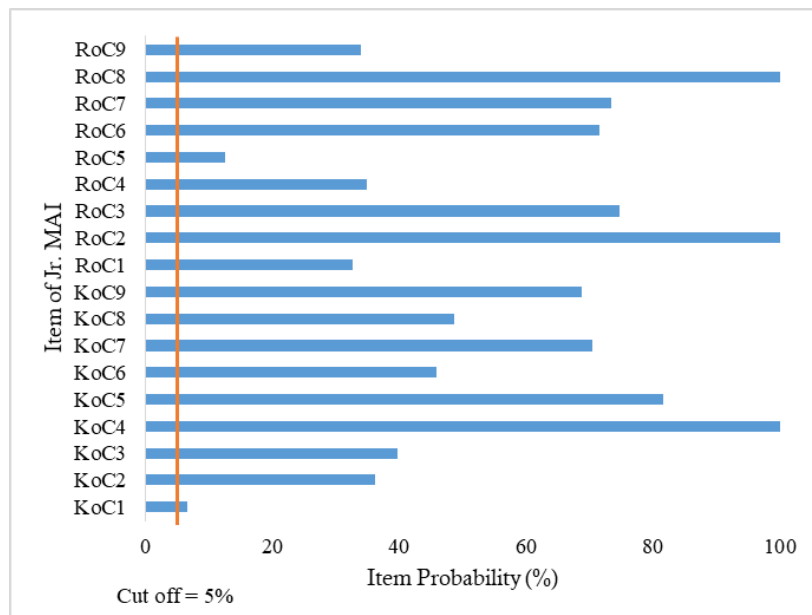

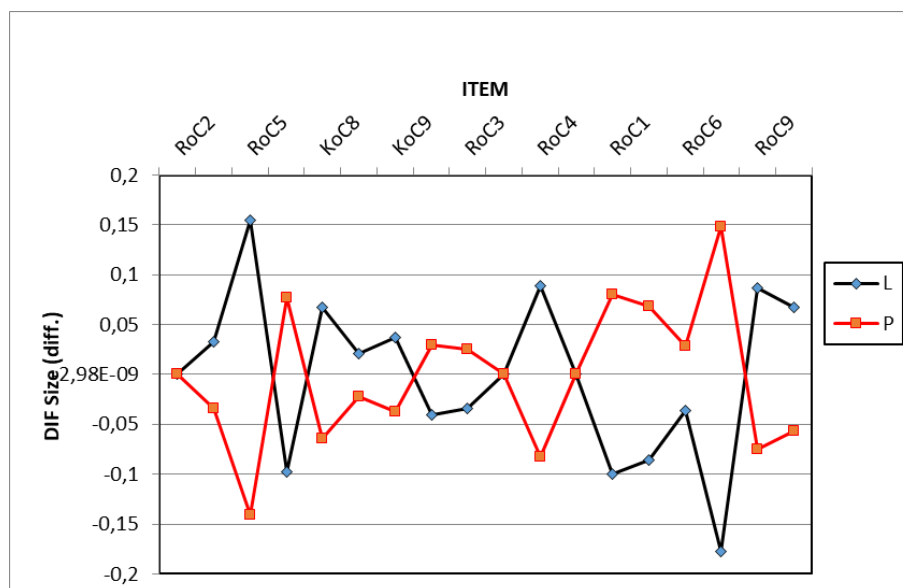**Figure 2.** Item probability across gender
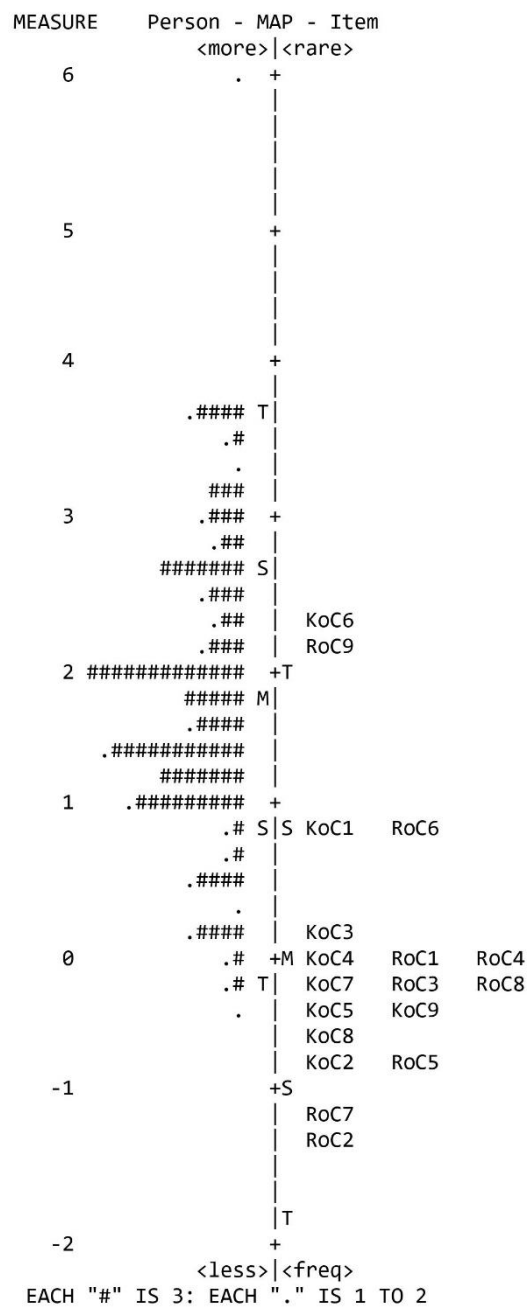

Figure 3. DIF size based on gender

Item and Person Distribution

The item and person relationship is depicted visually through the person-item map (Wright & Stone, 1979). Figure 4 shows the state of the person and item on the same logit scale. It is used to compare the difficulty

level of the item against the person's ability. The person-item map in Figure 4 is divided into 4 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 level of difficulty or students tend to easily agree with existing statements. The upper left area shows the position of the person with a high metacognitive level and the lower-left area shows the position of the person with a low metacognitive level. 18 items are distributed on the right side of the map. KoC6 item: "I know what the teacher expects me to learn" is the most difficult for students to agree on and RoC2 item: "I think about what I need to learn before I start working" is the easiest for students to agree on. There is a big gap between KoC3 and RoC9 items. Thus, to increase the sensitivity of Jr.MAI and reliability for person, some items need to be added (Muñoz & Nieto, 2019).

Person's mean score is higher than the item. This shows that the students' average chance of having metacognitive abilities is higher than the average item difficulty level. Items and persons that have the equivalent logit have a 50% probability of being agreed by the student. Items that are under a logit person have an agreed probability of more than 50%. Whereas items with a logit above person have a probability of less than 50% to be agreed on by students (Boone et al., 2014).

```
MEASURE     Person - MAP - Item
                <more>|<rare>
        6            .  +
                        |
                        |
                        |
                        |
                        |
        5               +
                        |
                        |
                        |
                        |
        4               +
                        |
               .#### T|
                 .#   |
                 .    |
                 ###  |
        3        .### +
                 .##  |
              ####### S|
                 .###  |
                  .##  |   KoC6
                 .###  |   RoC9
        2 ############# +T
                #####  M|
                 .#### |
           .########### |
                ####### |
        1    .######### +
                .# S|S KoC1      RoC6
                 .#   |
                 .#### |
                 .    |
                 .#### |   KoC3
        0         .#  +M KoC4      RoC1      RoC4
                 .# T|   KoC7      RoC3      RoC8
                 .   |   KoC5      KoC9
                     |   KoC8
                     |   KoC2      RoC5
       -1            +S
                     |   RoC7
                     |   RoC2
                     |
                     |
                     |T
       -2            +
                <less>|<freq>
        EACH "#" IS 3: EACH "." IS 1 TO 2
```

Figure 4. Person-item map of the Jr.MAI
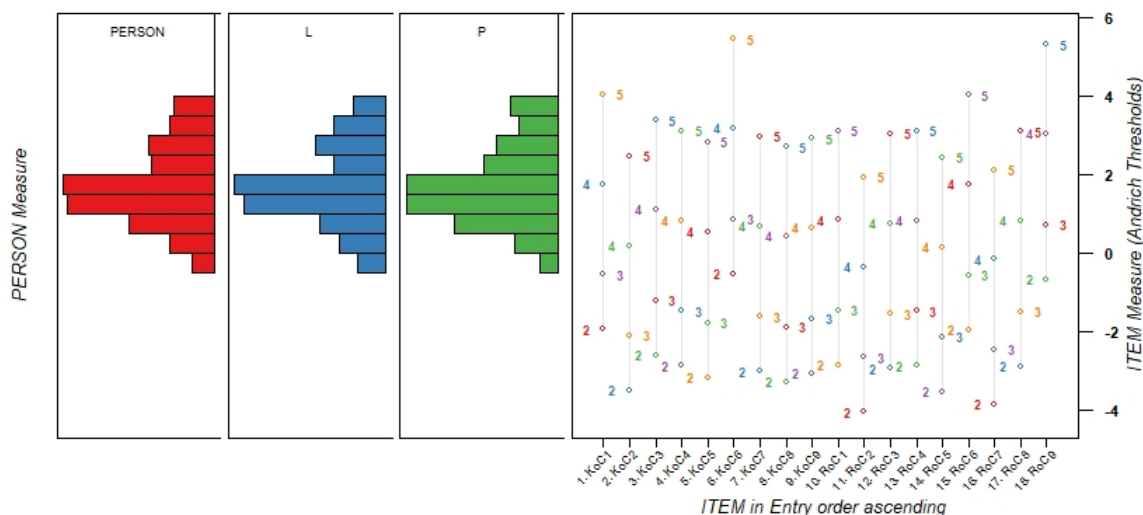
Figure 5. Person-item map based on Andrich threshold

To ensure the Jr.MAI in Indonesian context is fit well based using rating scales. We ran person item map for person and gender group based on Andrich Threshold in Figure 5. Figure 5 gives us further understanding of the distribution score in the questionnaire where all rating scales worked properly for all persons and gender in data fit and distribution.

CONCLUSION

Based on the results of the analysis and discussion previously presented, it was found that the use of the 5-point Likert scale in Jr. MAI is functioning well. Reliability Jr. MAI is in a good category and can classify items and people from more than three groups. The 18-item fit well against the model and was free of gender bias. So, Jr.MAI (Junior Metacognitive Awareness Inventory) has sufficient psychometric properties to measure the metacognitive abilities of high school students in Indonesia.

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 cultural attributes of the students 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 (Ning, 2018).

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).

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.

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, & Marais, I. (2019). *A course in rasch measurement theory: measuring in the educational, social and health sciences*. Springer Nature Singapore Pte Ltd.

Andrich, David. (2018). Advances in social measurement: a rasch measurement theory. *Perceived Health and Adaptation in Chronic Disease*, 66–91.

Aydin, U., & Ubuz, B. (2010). Turkish version of the junior metacognitive awareness inventory: an exploratory and confirmatory factor analysis. *Education and Science*, *35*(157), 32–47.

Bond, T., & Fox, C. (2015). *Applying the rasch model: fundamental measurement in the human sciences* (3rd ed.). Lawrence Erblaum & Associates.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Bradley, K. D., Peabody, M. R., Akers, K. S., & Knutson, N. (2015). Rating scales in survey research: using the rasch model to illustrate the middle category measurement flaw. *Survey Practice*, *8*(2).

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.

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.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

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), 34.

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), 1–25.

Jong, C., Hodges, T. E., Royal, K. D., & Welder, R. (2015). Instruments to measure elementary preservice teachers' conceptions: an application of the rasch rating scale model. *Educational Research Quarterly*, *39*(1), 21–48.

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.

Linacre, J. M. (1998). Detecting multidimensionality: which residual data-type works best?. *Journal of Outcome Measurement*, *2*, 266–283.

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: a rasch analysis. *European Journal of Preventive Cardiology*, 204748732090232.

Maryati, M., Prasetyo, Z. K., Wilujeng, I., & Sumintono, B. (2019). Measuring teachers' pedagogical content knowledge using many-facet rasch model. *Jurnal Cakrawala Pendidikan*, *38*(3), 452–464.

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). *Test adaptation standards for cross-lingual assessment.*

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*(JAN), 1–8.

Ning, H. K. (2017). The bifactor model of the junior metacognitive awareness inventory (Jr. MAI). *Current Psychology*, *38*(2), 367–375.

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.

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.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. In *Contemporary Educational Psychology* (Vol. 19, Issue 4, pp. 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.

Sriyanto, & Sukarelawan, M. I. (2019). Mapping of profile students' metacognitive awareness in yogyakarta, indonesia. *Jurnal Riset dan Kajian Pendidikan Fisika*, *6*(2), 56–62.

Sumintono, B, & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial*. Trim Komunikata Publishing House.

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.

Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies and academic achievement: pathways to achievement. *Metacognition and Learning*, *3*(2), 123–146.

Wright, B. D., & Stone, M. H. (1979). *Best test design: rasch measurement*. Mesa Press.

Wright, B., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.

Young, A., & Fry, J. D. (2008). Metacognitive awareness and academic achievement in college students. *Journal of the Scholarship of Teaching and Learning*, *8*(2), 1–10.

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.

Paper title:

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

Parts of review	Guidelines	Yes	Par tly	No	Reviewer's note for improvement	Author's respon (highlight of revision)
Title	• Does the subject matter fit within the scope of journal?	√				
	• Does the title clearly and sufficiently reflect its contents?	√				
Abstract	• Does the abstract contain informative, including Background, Methods, Results and Conclusion?	√				
Back-ground	• Is the background informative and sufficient (include the background problem and objectives)?	√				
	• Is research question of the study clear and understandable?	√				
	• Does the rationale of the study clearly explained using relevant literature?	√				
	• Is the "aim" of the manuscript clear and understandable?	√				
Methods	• Is the methodology chosen suitable to the nature of the topic studied?	√				
	• Is the methodology of the research described clearly?(including study design, location, subjects, data collection, data analysis)	√				
	• Is there adequate information about the data collection tools used? (only for empirical studies)	√				
	• Are the validity and reliability of data collection tools established? (only for empirical studies)	√				
	• Are the data collection tools suitable for the methodology of the study? (only for empirical studies)	√				
Results & Discussion	• Are the tables, graphs and pictures understandable, well presented and numbered consecutively?		√		The discussion is still not analyzed sharply.	
	• Do the data analysis and the interpretation appropriate to the problem and answer the objectives?		√			
	• Does the "discussion" section of the manuscript adequately relate to the current and relevant litarature?		√			
	• Are the findings discussed adequately considering the research question(s), sub-question(s) or hypothesis?		√		Partly	
Conclusion	• Is the conclusion clear and in the form of a narration instead of pointers?	√				
	• Isn't the conclusion a summary and consistent between problems, objectives and conclusion?		√			
References	• Do the references and citations match?	√				
	• Are the writing of references correct?			√		
Quality Criteria	• Do the title, problem, objectives, methods and conclusion are in line? Is it well organized?		√			
	• The quality of the language is satisfactory		√			
	• The work relevant and novel		√			
	• Are there strong consistencies among the parts of the manuscript? (introduction, methods, results and discussion, and conclusion)		√			

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

Abstract

Empirically, Metacognitive awareness is one of the main contributors to student 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 of Jr.MAI for high school students in Indonesian is still limited. The original Jr.MAI cannot be applied in Indonesia. Early in its development, the Jr.MAI self-report questionnaire was intended for students in the United States. Recently, there has been increasing interest in developing multilingual tests in the educational and psychological fields[MNAA1]. However, Studies were related to Jr.MAI in Indonesian context using Rasch measurement rarely conducted by researchers.[MNAA2] Therefore, this study aims to evaluate the psychometric properties of the Indonesian translation of the Junior Metacognitive Awareness Inventory (Jr.MAI) self-report questionnaire. The Jr.MAI questionnaire is consisted of 18 items and used a 5-point Likert scale response. 296 students (Male = 45.9% and Female = 54.1%) of public senior high schools in Indonesia completed the Jr.MAI questionnaire. The Rasch model had been used to evaluate the psychometric properties of Jr.MAI. The results showed that the 5-point rating scale with 18 items was functioning properly with good fit, no gender bias and achieving the unidimensionality and local independence assumptions, which proved that Jr. MAI questionnaire defined the latent variables and classified people and items well. Therefore, we concluded that The Jr.MAI questionnaire developed had good psychometric properties to be used by teachers and counselors for measuring and mapping the metacognitive characteristics at the senior high school level.

Keywords: rasch analysis; Jr.MAI; metacognitive

INTRODUCTION

Today, various literatures have reported the critical role of metacognition in the learning process of students[MNAA3][is4]. The ability to monitor and control learning has a positive correlation with learning success, increased academic achievement, and the health and well-being of students (Abdellah, 2015; Craig et al., 2020; Ning, 2018).[MNAA5][is6] For example, metacognitive strategies positively correlate with student test scores (Burin et al., 2020; Morphew, 2021). Also, the practical use of metacognitive skills and strategies has improved learning at various levels of education (Abdellah, 2015; Ahdhianto et al., 2020; Amin et al., 2020; Herlanti, 2015).

Metacognitive taxonomy has evolved in the last four decades. Flavell has become an 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 can be viewed as awareness or cognitive activity of a person about the thought process or everything related to it (Hidayat, Zulnaidi, & Zamri, 2018). Flavell (1979) proposed a metacognitive structure consisting of 4 main components (metacognitive knowledge, metacognitive experiences, goal, and action). About a decade and a half later, Schraw & Dennison proposed a metacognitive structure consisting of 2 main components: Metacognitive Knowledge (MK) and Metacognitive Experiences (ME). The MK component consists of three subcomponents, namely: declarative, procedural, and conditional knowledge. At the same time, the ME component consists of 5 sub-components, namely: 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 abilities (Sukarelawan & Sriyanto, 2019). Therefore, a standard instrument is needed to facilitate the task of teachers in assessing these abilities accurately. Several metacognitive questionnaires have been developed and applied in the field (Harrison & Vallin, 2018). This results from the existence of several metacognitive taxonomies that have been proposed previously (Allen & Armour-Thomas, 1993; Flavell, 1979; Schraw & Dennison, 1994). Because of their difficult nature to observe and assess, self-report questionnaires are the most effective and efficient and least problematic way to evaluate a person's metacognitive measures (Aydin & Ubuz, 2010; Craig et al., 2020).

The component proposed by Schraw & Dennison has produced numerous self-report questionnaires to evaluate metacognitive, 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 et al., 2015; Taasoobshirazi & Farley, 2013).

Sperling and colleague (2002) have developed Jr.MAI version B. Based on exploratory factor analysis, the 18-item Jr.MAI has concurrent validity and forms 2 metacognitive factors (Knowledge of Cognition and Regulation of Cognition). The research report shows that the 18-item model fits the two factors (see 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). Given that the Jr.MAI self-report instrument is intended to measure students' metacognition in the United States, 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 (Alindra et al., 2019; Bahari et al., 2020; Fauzi & Sa'diyah, 2019; Hidayat, Zulnaidi, & Syed Zamri, 2018; Sukarelawan & Sriyanto, 2019; Yasir et al., 2020). However, limited literature reports on the adaptation process and comprehensive study of the instrument's psychometric properties, especially the Jr. MAI version B. 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 when used in the Indonesian context.

Factor analysis techniques (exploratory factor analysis and confirmatory factor analysis) have been used to establish the construct validity of Jr.MAI version B (Aydin & Ubuz, 2010; Kim et al., 2017; Ning, 2019). Testing the rating scale function, unidimensionality, bias analysis through Differential Item Functioning (DIF) testing, item quality (item difficulty, and respondent's ability) are limitations that were not reported in previous studies (Aydin & Ubuz, 2010; Kim et al., 2017; Ning, 2019). So, we need a Rasch analysis technique based on the item response theory to fill this gap. As far as our observations, supported by Craig et al. (2020), there are limited reports on the Rasch analysis to evaluate the psychometric properties of Jr.MAI use in Indonesia. Therefore, this study aims to evaluate Jr. MAI's psychometric properties in the Indonesian context using the Rasch analysis technique. Thus, it is hoped that teachers or counselors can use Jr. MAI in measuring student metacognition in Indonesia.There have been no reports on Rasch analysis to evaluate the psychometric properties of Jr.MAI use in Indonesia. [MNAA7][is8]Therefore, this study aims to evaluate the psychometric properties of Jr.MAI in the Indonesian context using the Rasch analysis technique. Thus, it is hoped that

METHODS

Participants and procedures

Jr.MAI in Indonesia [MNAA9][is10]version was administered to several high schools in senior high school in Yogyakarta city, Indonesia. We chose Yogyakarta because it is the center of education and is identical to a "student city." Many students come from various regions in Indonesia to study here. So that researchers assume there is 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 of researchers and teachers, 351 students participated and filled the online questionnaire successfully. Students were selected using the convenience sampling technique. We ran data screening to exclude outliers before ~~doing~~ data analysis, 55 outliers were excluded from the dataset. Therefore, 296 students [MNAA11][is12](Male = 45.9% and Female = 54.1% were analyzed with Rasch measurement using Winsteps software. The number of participants in this analysis was adequate for data stability for sample size, more than 250 respondents (Chen et al., 2014; Lee et al., 2020).

Instrument

The Jr.MAI instrument (Sperling et al., 2002) has been adapted and translated into the Indonesian version by the Indonesian translator 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 each consist of 9 items (see 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 using Rasch measurement because Rasch measurement can solve some limitations of Classical Test Theory (CTT) analysis such as missing data in the analysis, reliability parameter only using Cronbach's alpha, the dependency of item and person which may not be reliable and valid in another research context (Rusch et al., 2017).

Data analysis

Data analysis began with data screening of participant responses. We applied the rating scale analysis using Rasch modelling to perform data analysis. The rating scale function was reviewed from the increase in the average observation, Andrich Threshold values, and probability curves (Andrich, 2018; Van Zile-Tamsen, 2017). Instrument reliability was determined based on the Cronbach alpha coefficient, person and item reliability parameter. 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

Screening Rating Scale

The psychometric properties of the 18-item Jr.MAI were analyzed using WINSTEPS 4.6.1 software. The analysis of the data starts with a person screening and rating. In the data screening process, 55 outliers have been detected. Outliers are a student whose answers are suspect and incompatible. Outfit MNSQ values are outside the acceptable criteria (0.5 to 1.5) indicated as misfitting person or outliers (Andrich, 2018; Bond et al., 2015). 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 an 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 a 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 ~~Statistical summary based on Rasch parameters~~

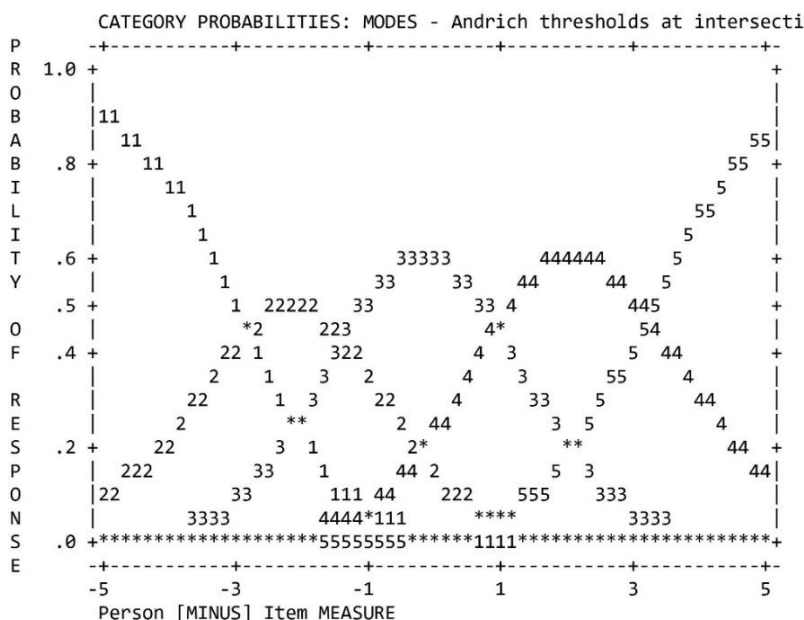|	Persons	Item
N	296	18
Mean	68.7	1131

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

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 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). All Categories on the probability curve have their respective peaks along the Measure axis. This 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. So the Jr.MAI rating scale used for Singapore 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.

Figure 1. Probability curves for the 5-point Likert scale of Jr.MAI

Table 1 showed that the mean of person ability is 1.74 logits, 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. The person separation proved that there are at least two person levels in this study, the student with high and low ability. Overall, the data have Chi-squared ($\chi2$) = 10072.06 (df= 10090), *p* > 0.05 indicating normal distribution achieved.[MNAA13]

Instrument Reliability
Screening Rating Scale
After person screening and rating, we calculated the summary statistics for the Jr.MAI questionnaire based on item and person parameters in Rasch modelling. Table 2 represents the Jr.MAI questionnaire's summary statistics for person and item based on Rasch parameters.

~~Analysis of the choice function on the provided rating scale is critical as an 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 2 shows a summary of the parameters used to assess the functioning of the options on the rating scale in Jr.MAI.~~

Table 2. Statistical summary based on Rasch parameters ~~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~~

	Persons	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 logits, 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. The person separation proved that there are at least two person levels in this study, the student with high and low ability. Overall, the data have Chi-squared (χ^2) = 10072.06 (df= 10090), $p > 0.05$ indicating normal distribution achieved.

Instrument reliability is estimated based on items and persons (see Table 1). The average person output is 1.74 logit. This indicates a tendency for respondents to agree on various attributes in Jr.MAI. Data items and persons were used to see the suitability of using items 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 special (Didino et al., 2019). This is supported by the value of the person and item reliability, respectively 0.83 and 0.99. These three reliability values indicate the items in Jr.MAI can define latent variables well (Maryati et al., 2019). The value of the person and item separation, represented in strata, is 3.31 and 13.24, respectively. This value indicates that Jr.MAI has a good ability to classify both person and item.

~~Based on Table 2, 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 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). All Categories on the probability curve have their respective peaks along the Measure axis. This 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. So the Jr.MAI rating scale used for Singapore 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.~~
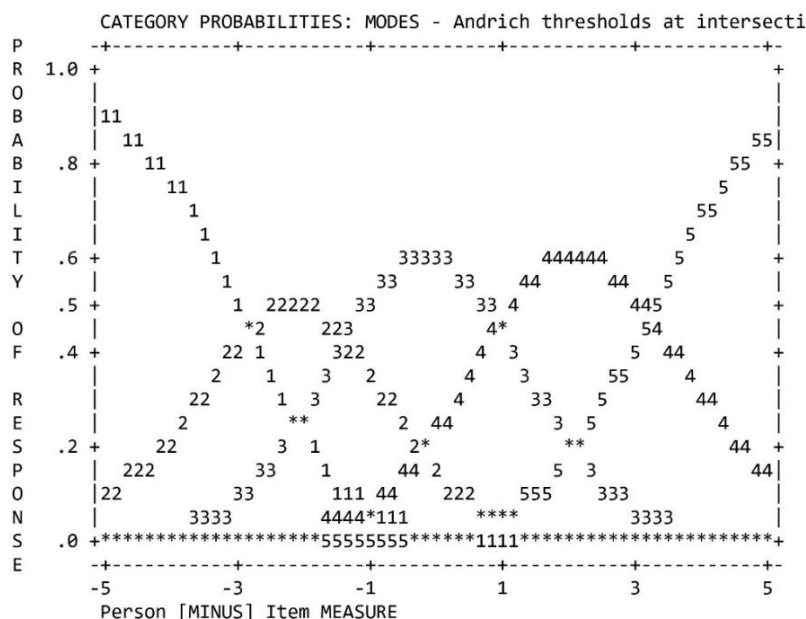
Figure 1. ~~Probability curves for the 5-point Likert scale of Jr.MAI~~

~~**Instrument Reliability**~~

~~Instrument reliability is estimated based on items and persons (see Table 1). The average person output is 1.74 logit. This indicates a tendency for respondents to agree on various attributes in Jr.MAI. Data items and persons were used to see the suitability of using items 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 special (Didino et al., 2019). This is supported by the value of the person and item reliability, respectively 0.83 and 0.99. These three reliability values indicate the items in Jr.MAI can define latent variables well (Maryati et al., 2019). The value of the person and item separation, represented in strata, is 3.31 and 13.24, respectively. This value indicates that Jr.MAI has a good ability to classify both person and item.~~

Unidimensionality and local independence

The unidimensionality of the Jr.MAI scale has been determined using the Principal Component Analysis of the residuals. The unidimensionality explains that the instrument is unidimension in measuring latent factor in this study, Metacognitive Awareness. Jr.MAI can achieve validity criteria in measuring latent factor or unidimension if the score of raw variance explained by measure is more than 30% (Chou & Wang, 2010). The value of raw variance explained by measures of the Jr.MAI questionnaire 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 item in the Jr.MAI questionnaire 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, which proves that the questionnaire is free of local dependence issues.

Item Fit

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). A fit item will make a good contribution in defining a common construct (Rahayu et al., 2020). In Rasch modelling, 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 (Andrich & Marais, 2019; Bond & Fox, 2015; Wright & Linacre, 1994). All 18 Jr.MAI items have infit values, and the MNSQ outfit is within the acceptance range. This shows a match in the response pattern to the target item and a match between person ability and item difficulty. Different from the findings reported by Ning. Two items (KoC6 and RoC1 have insufficient psychometric properties (see Table ~~2~~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 ~~2~~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 applied to measure the metacognitive measures of high school students in Indonesia.[MNAA14][is15]

Table 3. Comparison of infit and outfit MNSQ on Jr.MAI between Singapore students (Ning, 2018) and this study (Indonesian students).

Item	Statement	Indonesian students		Singapore students		PT-Mea~~sure~~ Corr.	Measure
		Infit MNSQ	Outfit MNSQ	Infit MNSQ	Outfit MNSQ		
KoC1	Saya tahu ketika saya memahami sesuatu.	0.85	0.85	0.96	0.92	0.57	0.82

KoC2	Ketika saya menyukai suatu topik, saya mempelajarinya lebih mendalam	0.95	0.94	1.18	1.18	0.58	-0.75
KoC3	Informasi-informasi penting dalam suatu topik sangat saya perhatikan.	0.79	0.79	0.90	0.93	0.60	0.15
KoC4	Saya dapat mengkondisikan diri saya untuk belajar ketika saya butuh.	1.04	1.07	0.80	0.86	0.54	-0.11
KoC5	Saya bisa belajar maksimal ketika sudah mengetahui sesuatu dari topik itu	1.23	1.23	1.17	1.18	0.53	-0.41
KoC6	Saya mengerti apa yang diharapkan guru pada saya untuk dipelajari	1.28	1.30	1.41	**1.55**	0.43	2.22
KoC7	Cara belajar yang berhasil saya gunakan sebelumnya, akan saya gunakan kembali	1.12	1.12	0.85	0.88	0.44	-0.25
KoC8	Saya menggunakan kekuatan-kekuatan belajar yang saya miliki untuk menutupi kelemahan-kelemahan.	0.81	0.80	0.78	0.80	0.67	-0.53
KoC9	Tanpa melalui proses berpikir, saya kadang-kadang langsung menggunakan strategi-strategi belajar.	0.97	0.97	1.17	1.33	0.50	-0.31
RoC1	Setelah menyelesaikan tugas sekolah, saya bertanya kepada diri sendiri apakah saya sudah mempelajari apa yang sebenarnya mau saya pelajari	1.10	1.11	1.53	**1.64**	0.50	-0.10
RoC2	Saya memikirkan apa yang sebenarnya ingin saya pelajari dari satu topik itu sebelum saya mulai mempelajarinya.	0.94	0.95	0.97	1.05	0.57	-1.29
RoC3	Ketika saya sedang mempelajari materi yang baru, saya bertanya pada diri sendiri seberapa baik saya melakukannya.	0.74	0.75	0.87	0.87	0.57	-0.19
RoC4	Ketika menyelesaikan tugas sekolah, saya mempertimbangkan beberapa cara penyelesaian lalu memilih cara terbaik	0.88	0.89	0.84	0.87	0.46	-0.11
RoC5	Setelah saya menyelesaikan tugas, saya bertanya pada diri sendiri apakah ada cara lain yang lebih mudah.	1.03	1.02	0.86	0.88	0.44	-0.79
RoC6	Untuk membantu memahami materi ketika belajar, saya membuat gambar-gambar atau diagram-diagram.	0.82	0.83	0.86	0.87	0.49	0.80
RoC7	Strategi-strategi belajar yang saya gunakan bisa saja berbeda tergantung pada tugasnya.	0.83	0.82	0.98	1.01	0.54	-1.10
RoC8	Sebelum mulai mengerjakan tugas, saya memastikan apa saja yang sudah harus dilakukan	1.24	1.22	1.14	1.14	0.54	-0.14
RoC9	Saya memeriksa kembali pekerjaan saya untuk memastikan bisa selesai tepat waktu	1.34	1.35	0.97	0.97	0.41	2.08

Differential Item Functioning (DIF)

One of the characteristics of a good instrument is if it does not have a bias towards certain respondent attributes. DIF analysis was conducted to see the trend of items in Jr.MAI on gender attributes. 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 < 5% as an indication of grain bias towards gender. We also ran a DIF analysis based on DIF size (see Figure 3). Jr.MAI proved that there is no gender bias on each item because no DIF size has |DIF| score ≥ 0.43 (slight to moderate) or |DIF| ≥ 0.64 logits (moderate to large) (Zwick et al., 1999). This result indicates the items in ~~Jr. MAI~~Jr.MAI are equal towards the male and female sex, which supports other findings from Papini et al. (2020). These findings are also in line with ~~Jr. MAI~~Jr.MAI in Singapore students (Ning, 2018).
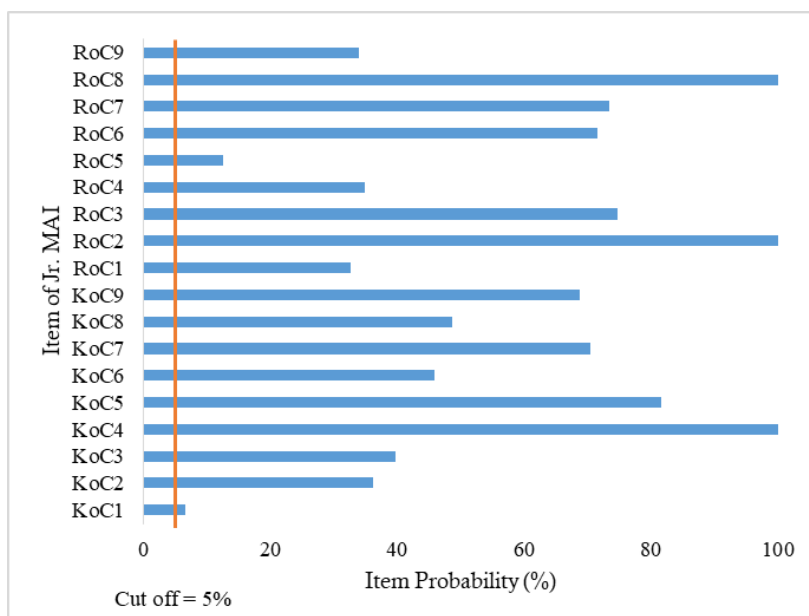
Figure 2. Item probability across gender

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 that there is no bias issue in all items.
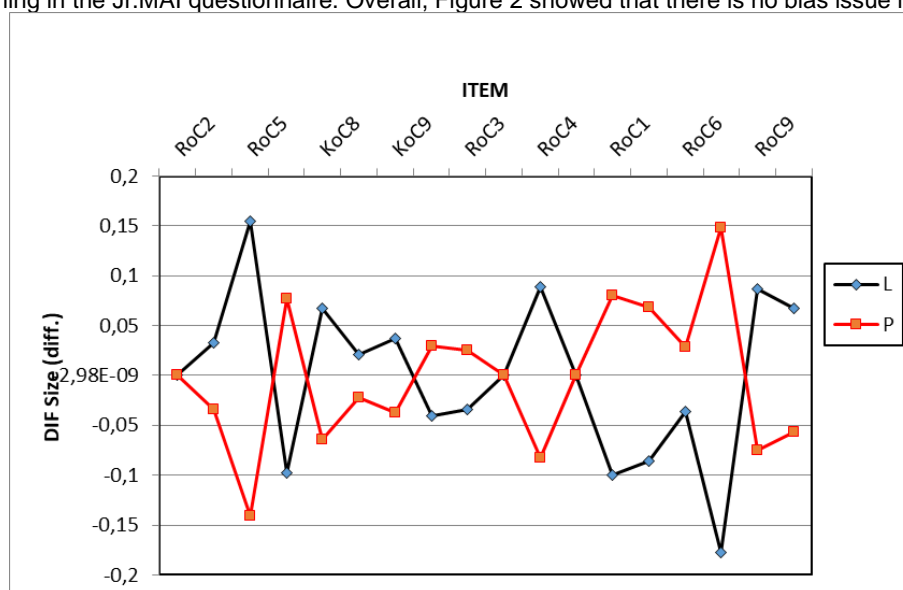

Figure 3. DIF size based on gender

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

Item and Person Distribution

The relationship between person and item is visualized through a person-item map (Wright map) (Wright & Stone, 1979). The Rasch model balances individual ability and item difficulty on a linear scale in one frame of reference so that person abilities and item difficulties are not interdependent (Blanc & Rojas, 2018; Wind & Gale, 2015). Figure 4 shows the state of the person and item on the same logit scale. It is used to compare the difficulty level of the item against the person's 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 level of difficulty, or students tend to agree with existing statements easily. The upper-left area shows the person's position with a high metacognitive level, and the lower-left area shows the person's position with a low metacognitive level. Eighteen items are distributed on the right side of the map. KoC6 item: "I know what the teacher expects me to learn" is the most difficult for students to agree on, and the RoC2 item: "I think about what I need to learn before I start working" the easiest for students to agree. There is a big gap between KoC3 and RoC9 items. Thus, to increase ~~Jr. MAI~~Jr.MAI's sensitivity and reliability, some items need to be added (Muñoz & Nieto, 2019).

A person's mean score is higher than the item. This shows that the students' average chance of having metacognitive abilities is higher than the average item difficulty level. Items and persons with the equivalent

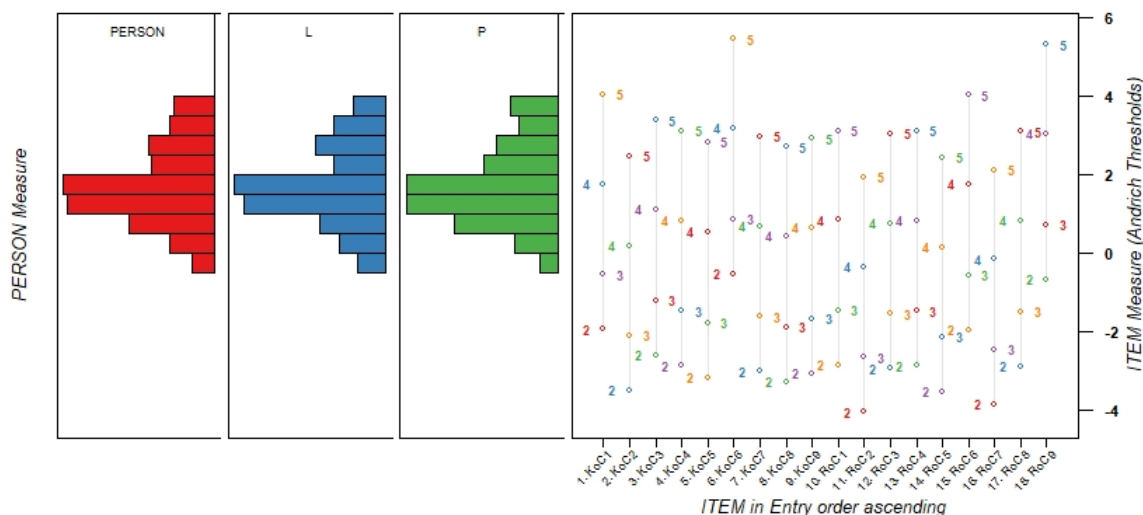logit have a 50% probability of being agreed by the student. Items that are under a logit person have an agreed probability of more than 50%. At the same time, items with a logit above person have a probability of less than 50% to be agreed on by students (Boone et al., 2014).

```
MEASURE     Person - MAP - Item
                <more>|<rare>
    6             .  +
                     |
                     |
                     |
                     |
                     |
    5                +
                     |
                     |
                     |
                     |
    4                +
                     |
            .#### T|
              .#   |
               .   |
             ###   |
    3       .###   +
              .##   |
          ####### S|
             .###   |
              .##   |  KoC6
              .###   |  RoC9
    2 #############  +T
           ##### M|
            .####   |
       .###########  |
           #######   |
    1    .#########  +
             .# S|S KoC1    RoC6
              .#   |
            .####   |
               .   |
            .####   |  KoC3
    0          .#  +M KoC4    RoC1    RoC4
              .# T|  KoC7    RoC3    RoC8
               .   |  KoC5    KoC9
                   |  KoC8
                   |  KoC2    RoC5
   -1                +S
                   |  RoC7
                   |  RoC2
                   |
                   |
                   |T
   -2                +
                <less>|<freq>
     EACH "#" IS 3: EACH "." IS 1 TO 2
```

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

The Person-item map indicated two items (RoC7 and RoC2) that easy to answers by students, but RoC7 and RoC2 are still below two standard deviation logit unit, so we can assume that these two items are no misbehave items. This result indicated that students' regulation of cognition, especially RoC7 and RoC2, is higher than other item domains. From the Jr.MAI person-item map, there is room for improvements of items by adding items that are more difficult to answer to cover all person abilities.

Figure 5. Person-item map based on Andrich threshold

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

CONCLUSION

Based on the results of the analysis and discussion previously presented, it was found that the use of the 5-point Likert scale in ~~Jr. MAI~~Jr.MAI is functioning well. Reliability ~~Jr. MAI~~Jr.MAI is in a good category and can classify items and people from more than three groups. The 18-item fit well against the model and was free of gender bias. So, Jr.MAI (Junior Metacognitive Awareness Inventory) has good psychometric properties to measure the metacognitive abilities of high school students in Indonesia. However, the Item-person map indicated that there is still room for improvement to cover all student ability by making some difficult items. Not substantial bias detected according to the probability of responses and DIF size based on 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 cultural attributes of the students 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, counsellors, and parents to help students achieve academic success (Ning, 2018).

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~~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).

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. https://doi.org/10.1016/j.sbspro.2015.01.707

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. https://doi.org/10.12973/eu-jer.9.2.753

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 (ICoIE 2018)*, *178*(ICoIE 2018), 318–322. https://doi.org/10.2991/icoie-18.2019.70

Allen, B. A., & Armour-Thomas, E. (1993). Construct Validation of Metacognition. *The Journal of Psychology*, *127*(2), 203–211. https://doi.org/10.1080/00223980.1993.9915555

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. https://doi.org/10.12973/eu-jer.9.1.143

Andrich, D. (2018). *Advances in Social Measurement: A Rasch Measurement Theory* (pp. 66–91). Routledge.

Andrich, D., & Marais, I. (2019). *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Science* (1st ed.). Springer. https://doi.org/10.1007/978-981-13-7496-8

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. *Proceedings of the Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019, 12 October 2019, Bandung, West Java, Indonesia*, *July*. https://doi.org/10.4108/eai.12-10-2019.2296408

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. https://doi.org/10.1371/journal.pone.0202551

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

Bond, T. G., Fox, C. M., & Lacey, H. (2007). Applying the Rasch model: Fundamental measurement. *In the Social Sciences (2nd*.

Boone, W. J., & Noltemeyer, A. (2017). Rasch Analysis: A Primer for School Psychology Researchers and Practitioners. *Cogent Education*, *4*(1), 1416898. https://doi.org/10.1080/2331186X.2017.1416898

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. In R. Glaser (Ed.), *Advances in Instructional Psychology* (Vol. 1, pp. 77–165). Erlbaum. http://www.eric.ed.gov/PDFS/ED146562.pdf.

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. https://doi.org/10.1007/s11409-020-09226-8

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. https://doi.org/10.1007/s11136-013-0487-5

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. https://doi.org/10.1177/0013164410379322

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. https://doi.org/10.1007/s11409-020-09222-y

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. https://doi.org/10.1186/s12955-019-1100-6

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. https://doi.org/10.2991/iccd-19.2019.8

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. https://doi.org/10.1007/s11409-017-9176-z

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., & Syed Zamri, S. N. A. (2018). Roles of metacognition and achievement goals in mathematical modeling competency: A structural equation modeling analysis. *PLOS ONE*, *13*(11), e0206211. https://doi.org/10.1371/journal.pone.0206211

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. https://doi.org/10.1371/journal.pone.0206211

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

Lee, W. L., 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

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. https://doi.org/10.21831/cp.v38i3.26598

Morphew, J. W. (2021). Changes in Metacognitive Monitoring Accuracy in an Introductory Physics Course. *Metacognition and Learning*, *16*(1), 89–111. https://doi.org/10.1007/s11409-020-09239-3

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. https://doi.org/10.3389/fpsyg.2019.00035

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

Ning, H. K. (2019). The Bifactor Model of the Junior Metacognitive Awareness Inventory (Jr. MAI). *Current Psychology*, *38*(2), 367–375. https://doi.org/10.1007/s12144-017-9619-3

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. https://doi.org/10.1177/1359105320904769

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. https://doi.org/10.1080/2331186X.2020.1721633

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. https://doi.org/10.1016/j.im.2016.06.005

Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

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. https://doi.org/10.1006/ceps.2001.1091

Sukarelawan, M. I., & Sriyanto, S. (2019). Mapping of Profile Students' Metacognitive Awareness in Yogyakarta, Indonesia. *Jurnal Riset Dan Kajian Pendidikan Fisika*, *6*(2), 56. https://doi.org/10.12928/jrkpf.v6i2.14556

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. https://doi.org/10.1080/09500693.2015.1104425

Taasoobshirazi, 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

Van Zile-Tamsen, C. (2017). Using Rasch Analysis to Inform Rating Scale Development. *Research in Higher Education*, *58*(8), 922–933. https://doi.org/10.1007/s11162-017-9448-0

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. https://doi.org/10.1002/sce.21172

Wright, B. D., & Stone, M. H. (1979). *Best Test Design: Rasch Measurement*. Mesa Press.

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. https://doi.org/10.22219/jpbi.v6i1.10081

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.

Paper title:

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

Parts of review	Guidelines	Yes	Partly	No	Reviewer's note for improvement	Author's responds (highlight of revision)
Title	• Does the subject matter fit within the scope of journal?	√				
	• Does the title clearly and sufficiently reflect its contents?	√				
Abstract	• Does the abstract contain informative, including Background, Methods, Results and Conclusion?	√				
Back-ground	• Is the background informative and sufficient (include the background problem and objectives)?	√				
	• Is research question of the study clear and understandable?	√				
	• Does the rationale of the study clearly explained using relevant literature?	√				
	• Is the "aim" of the manuscript clear and understandable?	√				
Methods	• Is the methodology chosen suitable to the nature of the topic studied?	√				
	• Is the methodology of the research described clearly?(including study design, location, subjects, data collection, data analysis)	√				
	• Is there adequate information about the data collection tools used? (only for empirical studies)	√				
	• Are the validity and reliability of data collection tools established? (only for empirical studies)	√				
	• Are the data collection tools suitable for the methodology of the study? (only for empirical studies)	√				
Results & Discussion	• Are the tables, graphs and pictures understandable, well presented and numbered consecutively?		√		The discussion is still not analyzed sharply.	We have elaborated the result and discussion section based on reviewer recommendations
	• Do the data analysis and the interpretation appropriate to the problem and answer the objectives?		√			
	• Does the "discussion" section of the manuscript adequately relate to the current and relevant litarature?		√			
	• Are the findings discussed adequately considering the research question(s), sub-question(s) or hypothesis?		√		Partly	We have elaborated the result and discussion section

						based on reviewer recommendations
Conclusion	• Is the conclusion clear and in the form of a narration instead of pointers?	√				
	• Isn't the conclusion a summary and consistent between problems, objectives and conclusion?		√			We have elaborated the conclusion section based on reviewer recommendations
References	• Do the references and citations match?	√				
	• Are the writing of references correct?			√		We have fixed and added the references based on reviewer recommendations. We have added up-to-date references from reputable journals. We have confirmed all the citations and the use of the APA style using the help of Mendeley's tools.
Quality Criteria	• Do the title, problem, objectives, methods and conclusion are in line? Is it well organized?		√			
	• The quality of the language is satisfactory		√			
	• The work relevant and novel		√			
	• Are there strong consistencies among the parts of the manuscript? (introduction, methods, results and discussion, and conclusion)		√			

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

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 high school students in Indonesian is still limited. The original Jr.MAI cannot be applied in Indonesia. By evaluating student metacognitive awareness using Jr.MAI, teachers can understand students' information and knowledge related to their learning strategies and learning behaviors. Therefore, this study aims to evaluate the psychometric properties of the Indonesian translation of the Junior Metacognitive Awareness Inventory (Jr.MAI) self-report questionnaire. The Jr.MAI questionnaire is consisted of 18 items and uses a 5-point Likert scale response. 296 students (Male = 45.9% and Female = 54.1%) of public senior high schools in Indonesia completed the Jr.MAI questionnaire. The Rasch model had been used to evaluate the psychometric properties of Jr.MAI. The results showed that the 5-point rating scale with 18 items functioned properly with a good fit, no gender bias, and achieving the unidimensionality and local independence assumptions, proving that the Jr. MAI questionnaire defined the latent variables and classified people and items well. Therefore, we concluded that The Jr.MAI questionnaire developed had good psychometric properties to be used by teachers and counselors for measuring and mapping the metacognitive characteristics at the senior high school level.

Keywords: Rasch analysis; Jr.MAI; metacognitive

INTRODUCTION

Various kinds of literature have reported 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 the health and well-being of students (Abdellah, 2015; Craig et al., 2020; Ning, 2018). For example, metacognitive strategies positively correlate with students' test scores (Burin et al., 2020; Morphew, 2021). Also, the practical use of metacognitive skills and strategies has improved learning at various levels of education (Abdellah, 2015; Ahdhianto et al., 2020; Amin et al., 2020; Herlanti, 2015).

Metacognitive taxonomy has evolved in the last four decades. Flavell has become an 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 can be viewed as a person's awareness or cognitive activity about the thought process or everything related to it (Hidayat, Zulnaidi, & Zamri, 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 consisting of 2 main components: Metacognitive Knowledge (MK) and Metacognitive Experiences (ME). The MK component consists of three subcomponents, namely: 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 abilities (Sukarelawan & Sriyanto, 2019). Therefore, a standard instrument is needed to facilitate the task of teachers in assessing these abilities accurately. Several metacognitive questionnaires have been developed and applied in the field (Harrison & Vallin, 2018). This statement results from the existence of several metacognitive taxonomies that have been proposed previously (Allen & Armour-Thomas, 1993; Flavell, 1979; Schraw & Dennison, 1994). Because of their complex nature to observe and assess, self-report questionnaires are the most effective, efficient, and least problematic way to evaluate a person's metacognitive measures (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 et al., 2015; Taasoobshirazi & Farley, 2013).

Sperling and colleagues (2002) have developed Jr.MAI version B. The 18-item Jr.MAI has concurrent validity and forms 2 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 (see 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 (Alindra et al., 2019; Bahari et al., 2020; Fauzi & Sa'diyah, 2019; Hidayat, Zulnaidi, & Syed Zamri, 2018; Sukarelawan & Sriyanto, 2019; Yasir et al., 2020). However, limited literature reports on the adaptation process and comprehensive study of the instrument's psychometric properties, especially the Jr. MAI version B. 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 when used in the Indonesian context.

Factor analysis techniques (exploratory factor analysis and confirmatory factor analysis) have been 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. As far as 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

Participants and procedures

Jr.MAI in Indonesia version was administered to several high schools in senior high school in Yogyakarta, Indonesia. The researchers chose Yogyakarta because it is the center of education and is identical to a "student city." Many students come from various regions in Indonesia to study here, 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 of researchers and teachers, 351 students

participated and filled 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% and Female = 54.1%) were analyzed with Rasch measurement using Winsteps software. The number of participants in this analysis was adequate for data stability for sample size, more than 250 respondents (Chen et al., 2014; Lee et al., 2020).

Instrument

The Jr.MAI instrument (Sperling et al., 2002) has been adapted and translated into the Indonesian version by the Indonesian translator 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 each consist of 9 items (see 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 using 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, the dependency of item and person, which may not be reliable and valid in another research context (Rusch et al., 2017).

Data analysis

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, 2018; Van Zile-Tamsen, 2017). Instrument reliability was determined based on the Cronbach alpha coefficient, person and item reliability parameter, and 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

Screening Rating Scale
The psychometric properties of the 18-item Jr.MAI were analyzed using WINSTEPS 4.6.1 software. The analysis of the data starts with a person screening and rating. In the data screening process, 55 outliers have been detected. Outliers are students with suspicious and incompatible answers. Outfit MNSQ values are outside the acceptable criteria (0.5 to 1.5) indicated as misfitting persons or outliers (Andrich, 2018; Bond et al., 2015). 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 an 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 a 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). 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. So 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.
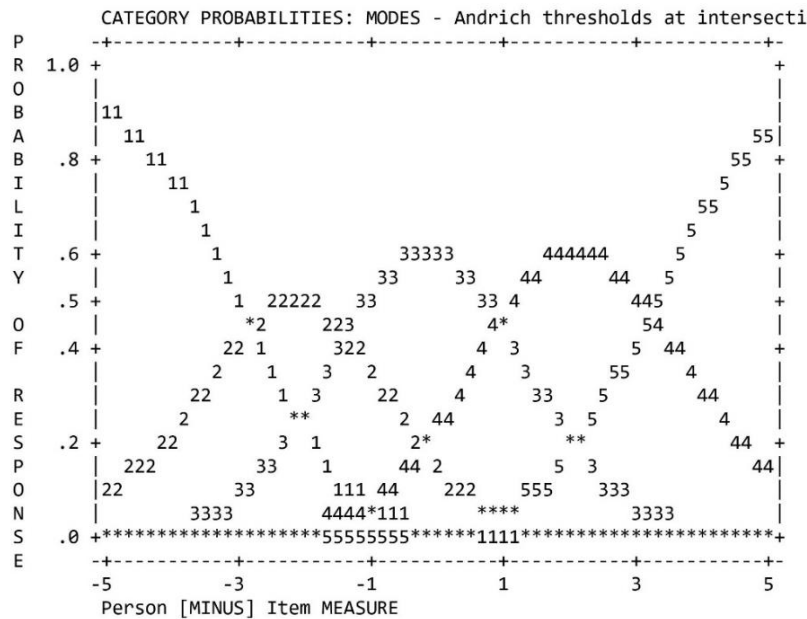
Figure 1. Probability curves for the 5-point Likert scale of Jr.MAI

Instrument Reliability
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 summary statistics for person and item based on Rasch parameters.

Table 2. Statistical summary based on Rasch parameters

	Persons	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 logits, 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 (χ^2) = 10072.06 (df= 10090), $p > 0.05$ indicating normal distribution achieved.

Instrument reliability is estimated based on items and persons (see Table 2). The average person output is 1.74 logit, indicating a tendency for respondents to agree on various attributes in Jr.MAI. Data items and persons were used to see the suitability of using items 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 has an excellent ability to classify both person and item.

Unidimensionality and local independence
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 unidimensionality if the score of raw variance explained by measure is more than 30% (Chou & Wang, 2010). The value of raw variance explained by measures of the Jr.MAI questionnaire 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 item in the Jr.MAI questionnaire 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.

Item Fit

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). 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 (Andrich & Marais, 2019; Bond & Fox, 2015; Wright & Linacre, 1994). 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 (see 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 applied to measure the metacognitive measures of high school students in Indonesia.

Table 3. Comparison of infit and outfit MNSQ on Jr.MAI between Singapore students (Ning, 2018) and this study (Indonesian students).

Item	Statement	Indonesian students		Singapore students		PT-Mea. Corr.	Measure
		Infit MNSQ	Outfit MNSQ	Infit MNSQ	Outfit MNSQ		
KoC1	*Saya tahu ketika saya memahami sesuatu.*	0.85	0.85	0.96	0.92	0.57	0.82
KoC2	*Ketika saya menyukai suatu topik, saya mempelajarinya lebih mendalam*	0.95	0.94	1.18	1.18	0.58	-0.75
KoC3	*Informasi-informasi penting dalam suatu topik sangat saya perhatikan.*	0.79	0.79	0.90	0.93	0.60	0.15
KoC4	*Saya dapat mengkondisikan diri saya untuk belajar ketika saya butuh.*	1.04	1.07	0.80	0.86	0.54	-0.11
KoC5	*Saya bisa belajar maksimal ketika sudah mengetahui sesuatu dari topik itu*	1.23	1.23	1.17	1.18	0.53	-0.41
KoC6	*Saya mengerti apa yang diharapkan guru pada saya untuk dipelajari*	1.28	1.30	1.41	**1.55**	0.43	2.22
KoC7	*Cara belajar yang berhasil saya gunakan sebelumnya, akan saya gunakan kembali*	1.12	1.12	0.85	0.88	0.44	-0.25
KoC8	*Saya menggunakan kekuatan-kekuatan belajar yang saya miliki untuk menutupi kelemahan-kelemahan.*	0.81	0.80	0.78	0.80	0.67	-0.53
KoC9	*Tanpa melalui proses berpikir, saya kadang-kadang langsung menggunakan strategi-strategi belajar.*	0.97	0.97	1.17	1.33	0.50	-0.31
RoC1	*Setelah menyelesaikan tugas sekolah, saya bertanya kepada diri sendiri apakah saya sudah mempelajari apa yang sebenarnya mau saya pelajari*	1.10	1.11	1.53	**1.64**	0.50	-0.10
RoC2	*Saya memikirkan apa yang sebenarnya ingin saya pelajari dari satu topik itu sebelum saya mulai mempelajarinya.*	0.94	0.95	0.97	1.05	0.57	-1.29
RoC3	*Ketika saya sedang mempelajari materi yang baru, saya bertanya pada diri sendiri seberapa baik saya melakukannya.*	0.74	0.75	0.87	0.87	0.57	-0.19
RoC4	*Ketika menyelesaikan tugas sekolah, saya mempertimbangkan beberapa cara penyelesaian lalu memilih cara terbaik*	0.88	0.89	0.84	0.87	0.46	-0.11
RoC5	*Setelah saya menyelesaikan tugas, saya bertanya pada diri sendiri apakah ada cara lain yang lebih mudah.*	1.03	1.02	0.86	0.88	0.44	-0.79
RoC6	*Untuk membantu memahami materi ketika belajar, saya membuat gambar-gambar atau diagram-diagram.*	0.82	0.83	0.86	0.87	0.49	0.80
RoC7	*Strategi-strategi belajar yang saya gunakan bisa saja berbeda tergantung pada tugasnya.*	0.83	0.82	0.98	1.01	0.54	-1.10

| RoC8 | Sebelum mulai mengerjakan tugas, saya memastikan apa saja yang sudah harus dilakukan | 1.24 | 1.22 | 1.14 | 1.14 | 0.54 | -0.14 |
| RoC9 | Saya memeriksa kembali pekerjaan saya untuk memastikan bisa selesai tepat waktu | 1.34 | 1.35 | 0.97 | 0.97 | 0.41 | 2.08 |

Differential Item Functioning (DIF)

One of the characteristics of a good instrument is if it does not bias specific respondent attributes. DIF analysis was conducted to see the trend of items in Jr.MAI on gender attributes. 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 < 5% indicating grain bias towards gender. We also ran a DIF analysis based on DIF size (see Figure 3). Jr.MAI proved that there is no gender bias on each item because no DIF size has |DIF| score ≥ of 0.43 (slight to moderate) or |DIF| ≥ 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 from Papini et al. (2020). These findings are also in line with Jr.MAI in Singaporean students (Ning, 2018).
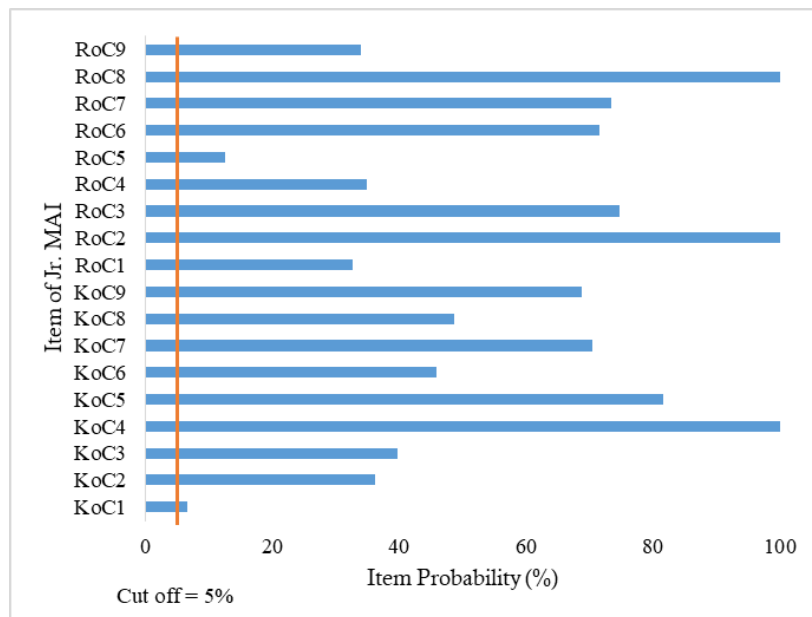

Figure 2. Item probability across gender

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 that there is no bias issue in all items.
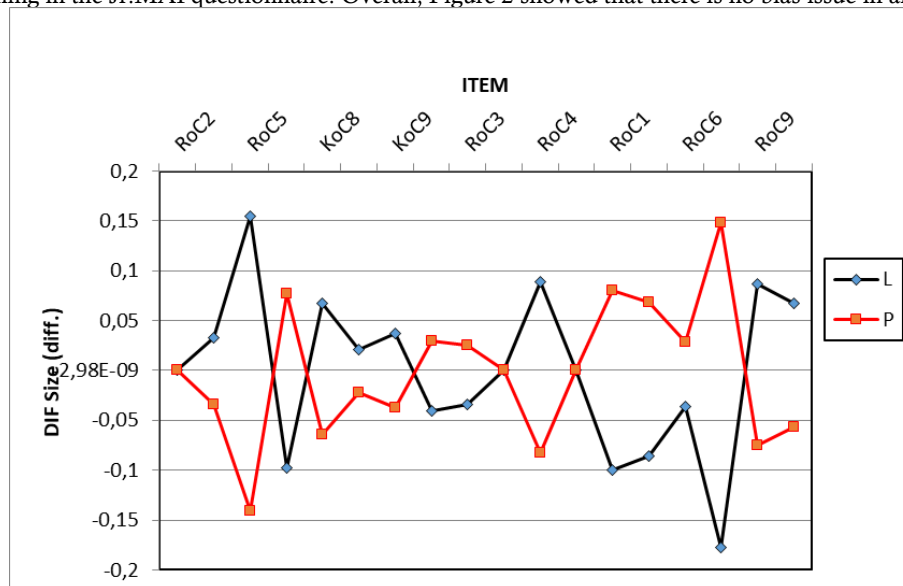

Figure 3. DIF size based on gender

We also performed DIF analysis based on gender (Female (P) and Male (L)). 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, which is less than 0.43 (slight to moderate categories).

Item and Person Distribution

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 one frame of reference so that person abilities and item difficulties are not interdependent (Blanc & Rojas, 2018; Wind & Gale, 2015). Figure 4 shows the state of the person and item on the same logit scale. It is used to compare the difficulty level of the item 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 person's position with a high metacognitive level, and the lower-left area shows the person's position with a low metacognitive level. Eighteen items are distributed on the right side of the map. KoC6 item: "*Saya mengerti apa yang diharapkan guru pada saya untuk dipelajari.*" is the most difficult for students to agree on, and the RoC2 item: "*Saya memikirkan apa yang sebenarnya ingin saya pelajari dari satu topik itu sebelum saya mulai mempelajarinya.*" 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).

A person's mean score is higher than the item. It shows that the students' average chance of having metacognitive abilities 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 that are 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).
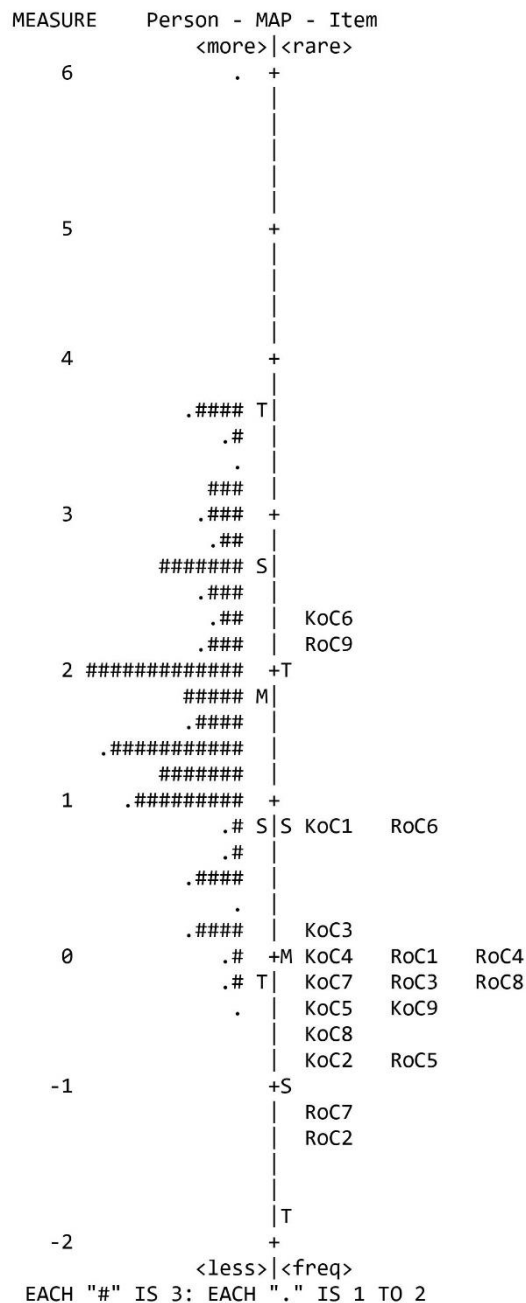
```
MEASURE     Person - MAP - Item
                  <more>|<rare>
   6               .  +
                      |
                      |
                      |
                      |
   5                  +
                      |
                      |
                      |
                      |
   4                  +
                      |
             .#### T|
               .#   |
               .    |
              ###   |
   3         .###   +
              .##   |
           ####### S|
              .###  |
              .##   |   KoC6
              .###  |   RoC9
   2 ############   +T
            ##### M|
             .#### |
          .########### |
            ####### |
   1    .######### +
              .# S|S KoC1      RoC6
              .#   |
            .####  |
              .    |
            .####  |   KoC3
   0          .#  +M KoC4    RoC1    RoC4
              .# T|   KoC7    RoC3    RoC8
              .   |   KoC5    KoC9
                  |   KoC8
                  |   KoC2    RoC5
  -1              +S
                  |   RoC7
                  |   RoC2
                  |
                  |
                  |T
  -2              +
               <less>|<freq>
     EACH "#" IS 3: EACH "." IS 1 TO 2
```

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

The Person-item map indicated two items (RoC7 and RoC2) that students easily answer, but RoC7 and RoC2 are still below two standard deviation logit units, so we can assume that these two items are no misbehave items. This result indicated that students' regulation of cognition, especially RoC7 and RoC2, is higher than other item domains. From the Jr.MAI person-item map, there is room for improvements of items by adding items that are more difficult to answer to cover all person abilities.
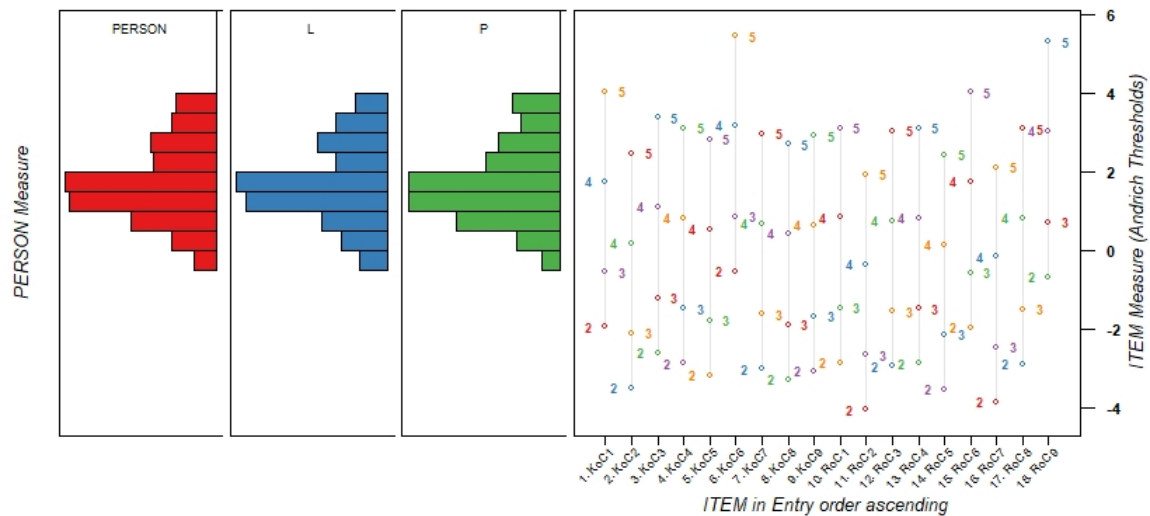
Figure 5. Person-item map based on Andrich threshold

To ensure the Jr.MAI in the Indonesian context is fit well based using rating scales, we ran a person-item map for the person and gender groups based on Andrich Threshold in Figure 5. Figure 5 gives us a further understanding of the distribution score in the questionnaire where all rating scales worked adequately for all persons 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. Reliability Jr.MAI is in a good category and can classify items and people from more than three groups. The 18-item fit well against the model and was free of gender bias. So, Jr.MAI (Junior Metacognitive Awareness Inventory) has good psychometric properties to measure the metacognitive abilities of high school students in Indonesia. However, the Item-person map indicated that there is still room for improvement to cover all students' abilities by making some difficult items. Not substantial bias was detected according to the probability of responses and DIF size based on 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 cultural attributes of the students 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 (Ning, 2018).

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).

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. https://doi.org/10.1016/j.sbspro.2015.01.707

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. https://doi.org/10.12973/eu-jer.9.2.753

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 (ICoIE 2018)*, *178*(ICoIE 2018), 318–322. https://doi.org/10.2991/icoie-18.2019.70

Allen, B. A., & Armour-Thomas, E. (1993). Construct Validation of Metacognition. *The Journal of Psychology*, *127*(2), 203–211. https://doi.org/10.1080/00223980.1993.9915555

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. https://doi.org/10.12973/eu-

jer.9.1.143

Andrich, D. (2018). *Advances in Social Measurement: A Rasch Measurement Theory* (pp. 66–91). Routledge.

Andrich, D., & Marais, I. (2019). *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Science* (1st ed.). Springer. https://doi.org/10.1007/978-981-13-7496-8

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. *Proceedings of the Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019, 12 October 2019, Bandung, West Java, Indonesia, July*. https://doi.org/10.4108/eai.12-10-2019.2296408

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. https://doi.org/10.1371/journal.pone.0202551

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

Bond, T. G., Fox, C. M., & Lacey, H. (2007). Applying the Rasch model: Fundamental measurement. *In the Social Sciences (2nd*.

Boone, W. J., & Noltemeyer, A. (2017). Rasch Analysis: A Primer for School Psychology Researchers and Practitioners. *Cogent Education*, *4*(1), 1416898. https://doi.org/10.1080/2331186X.2017.1416898

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. In R. Glaser (Ed.), *Advances in Instructional Psychology* (Vol. 1, pp. 77–165). Erlbaum. http://www.eric.ed.gov/PDFS/ED146562.pdf.

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. https://doi.org/10.1007/s11409-020-09226-8

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. https://doi.org/10.1007/s11136-013-0487-5

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. https://doi.org/10.1177/0013164410379322

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. https://doi.org/10.1007/s11409-020-09222-y

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. https://doi.org/10.1186/s12955-019-1100-6

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. https://doi.org/10.2991/iccd-19.2019.8

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. https://doi.org/10.1007/s11409-017-9176-z

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., & Syed Zamri, S. N. A. (2018). Roles of metacognition and achievement goals in mathematical modeling competency: A structural equation modeling analysis. *PLOS ONE*, *13*(11), e0206211. https://doi.org/10.1371/journal.pone.0206211

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. https://doi.org/10.1371/journal.pone.0206211

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

Lee, W. L., 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

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. https://doi.org/10.21831/cp.v38i3.26598

Morphew, J. W. (2021). Changes in Metacognitive Monitoring Accuracy in an Introductory Physics Course. *Metacognition and Learning*, *16*(1), 89–111. https://doi.org/10.1007/s11409-020-09239-3

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. https://doi.org/10.3389/fpsyg.2019.00035

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

Ning, H. K. (2019). The Bifactor Model of the Junior Metacognitive Awareness Inventory (Jr. MAI). *Current Psychology*, *38*(2), 367–375. https://doi.org/10.1007/s12144-017-9619-3

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. https://doi.org/10.1177/1359105320904769

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. https://doi.org/10.1080/2331186X.2020.1721633

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. https://doi.org/10.1016/j.im.2016.06.005

Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

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. https://doi.org/10.1006/ceps.2001.1091

Sukarelawan, M. I., & Sriyanto, S. (2019). Mapping of Profile Students' Metacognitive Awareness in Yogyakarta, Indonesia. *Jurnal Riset Dan Kajian Pendidikan Fisika*, *6*(2), 56. https://doi.org/10.12928/jrkpf.v6i2.14556

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. https://doi.org/10.1080/09500693.2015.1104425

Taasoobshirazi, 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

Van Zile-Tamsen, C. (2017). Using Rasch Analysis to Inform Rating Scale Development. *Research in Higher Education*, *58*(8), 922–933. https://doi.org/10.1007/s11162-017-9448-0

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. https://doi.org/10.1002/sce.21172

Wright, B. D., & Stone, M. H. (1979). *Best Test Design: Rasch Measurement*. Mesa Press.

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. https://doi.org/10.22219/jpbi.v6i1.10081

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.