The metacognitive awareness of reading strategy among pre-service primary teachers and the possibility of rating improvement using Rasch analysis☆

Soeharto Soeharto [a,b,*,1], Martono Martono [c,2], Hairida Hairida [c,3], Aigul Akhmetova [b,4], Fitria Arifiyanti [b,5], Csapó Benő [d,e,6], Charalambous Charalambos [f,7]

[a] Research Center of Educational Technologies, Azerbaijan State University of Economics, Baku, Azerbaijan
[b] Doctoral School of Education, University of Szeged, Szeged, Hungary
[c] Faculty of teacher training and education, Tanjungpura University, Indonesia
[d] Institute of Education, University of Szeged, Hungary
[e] MTA-SZTE Research Group on the Development of Competencies, Hungary
[f] Department of Education, University of Cyprus, Cyprus

ARTICLE INFO

ABSTRACT

This study investigated pre-service primary teachers' reading strategies using the Metacognitive Awareness of Reading Strategy Inventory (MARSI) in the Indonesian version. The participants were 4377 pre-service primary teachers in Indonesia. The data were collected via an online form, and the invitations were sent out randomly. The participants were divided into two studies: 2912 participants were in study 1, which used a 5-point Likert scale, and 1465 participants were in study 2, using a 4-point Likert scale. The Rasch measurement approach was used in investigating the validity and reliability, rating improvement analysis, checking item difficulty functioning based on gender, and classifying pre-service primary teacher reading strategies. The adapted questionnaire satisfied all psychometric requirements based on Rasch modeling. The questionnaire with a 4-point scale showed better function than the 5-point one based on the observed measure average and curve probability category. The teachers' abilities ranged from − 6 logit to 6.45 logit.

1. Introduction

Reading is a crucial skill for teaching and learning in all academic domains. Developing reading strategies is the basic foundation for further literacy growth (Cunningham & Stanovich, 2001; Wigfield et al., 2016). Practical improvements in reading skills are possible if the reader or learner is using efficient and useful reading strategies, which produce valid "interpretation of the meanings communicated in the text" (Van Gelderen et al., 2004, p. 19). However, very little research has been done on reading strategies in pre-service primary teachers, how they are to teach and observe reading improvements in young learners, or how this process should be managed and evaluated further in the learning process. Several researchers (e.g., Wilson & Bai, 2010; Yorulmaz et al., 2021) indicate that pre-service primary teachers should be proficient in reading skills to teach students.

No assessment for reading skills or reading strategies in higher

education has been developed for the Indonesian context (Sarimanah et al., 2022). Previous studies have focused on assessing students' reading skills and learning achievements (Deliany & Cahyono, 2020a; Fitrisia et al., 2015; Halim et al., 2022; Sarimanah et al., 2022). Fitrisia et al. (2015) investigated English reading comprehension tests and metacognitive awareness of reading strategies at the senior high school level, but that study did not employ objective measurements. In addition, it was a replication study that did not develop new techniques or findings, such as a detailed analysis of rating improvements and personal abilities. Halim et al. (2022) studied the relationship between metacognitive awareness of reading strategies and reading comprehension at the university level. However, their sample size was small (191 respondents) and was only drawn from a private university in Indonesia, limiting its generalizability. In this study, we thoroughly investigated metacognitive awareness of reading strategies at the university level among pre-service primary teachers in Indonesia. Instrument scaling and rating improvements were performed to identify the validation and suitability of the instrument when used in different contexts.

1.1. Metacognitive awareness of reading strategies

Reading is the process of decoding written symbols (see, e.g., Cunningham & Stanovich, 2001; Wigfield et al., 2016) and is a complex "interactive and cognitive process of meaning construction" (Hog-Nam & Page, 2014, p. 195). Reading is an essential skill, as "proficient reading comprehension is crucial for success in every academic domain" (Wigfield et al., 2016, p. 190), and is pivotal for utilizing the meaning of reading materials (Mokhtari & Reichard, 2004; Mahdavi & Tensfeldt, 2013).

Recent studies have determined (Wilson, & Bai, 2010; Akhmetova et al., 2022) that more experienced and sophisticated readers use more reading strategies in practice and have greater mastery of them. The literature also reports that readers who read a great deal can adopt more reading strategies and skills than those who do not read frequently, as proficient readers are constantly showing mastery over metacognitive reading awareness (as cited in Akhmetova et al., 2022, p. 2). Metacognitive reading awareness is a decisive approach to communicating and understanding one's own motivation, beliefs, suggestions, attitude, etc. (Mokhtari & Reichard, 2002; Sheorey & Mokhtari, 2001)

1.2. Metacognitive awareness of reading strategy in higher education

Reading strategies should be appropriately developed and utilized. Previous studies show that readers' reading ability improves with the proficient use of appropriate strategies. Likewise, scholars and researchers (Sulentic-Dowell et al., 2006; Wilson & Bai, 2010) have found that readers, teachers, and/or pre-service teachers should have solid and well-determined language skills with broad and rich lexicons. This is necessary for strategic reading awareness and teaching comprehension.

Researchers, educators, and scholars (Clark & Graves, 2005; Pressley, 2008) have found that experienced and more proficient readers use more developed reading strategies and closely link their knowledge to metacognitive reading awareness. Likewise, in higher education, reading processes require constant practice and training for teachers, learners, and pre-service teachers, who should be ready to instruct students and provide them with guidelines for better understanding reading texts (see, e.g., Clark & Graves, 2005; Pressley, 2008; Wilson & Bai, 2010). Some research studies have claimed, "in practice, many teachers lack sufficient knowledge about metacognition" (Veenman, van Hout-Wolters & Afflerbac, 2006 as cited in Wilson & Bai, 2010, p. 272) and may have a poor understanding of the frequency and necessity of the use of instructional strategies in the reading process.

Other researchers find that teachers should have frequent reading habits and show a good "reading model" to their learners while using "good and effective reading strategies," and teachers should be focused

on students' ultimate success and achievements (Akyol & Ulusoy, 2010, pp. 878–879). Some researchers (Hall, 2005; Lesley et al., 2007) have found that pre-and in-service teachers themselves often think that they are not well-prepared to teach students reading skills due to their own limited knowledge in reading, their former "negative attitudes towards the act of reading" (p. 879) and/or the imperfect connection of the reading materials to the world (Akyol & Ulusoy, 2010). In addition, several studies (Akyol & Ulusoy, 2010; Conaway et al., 2003) show that when pre-and in-service teachers do not frequently use reading strategies in their teaching process about half of the time because they concentrate on the content, not on reading as subject; little information has been obtained regarding the effectiveness of their teaching process or possible improvements to their educational programs.

The MARSI questionnaire is a tool frequently used in research to gauge individuals' awareness and application of metacognitive reading strategies. This questionnaire consists of multiple statements that ask participants to indicate how often they use specific strategies while reading academic or educational materials (Deliany & Cahyono, 2020; Levchyk et al., 2022). However, the study by Deliany and Cahyono (2020), which focused on assessing EFL students in an Indonesian context, had a significant shortcoming due to the absence of a valid and reliable analysis. Another Indonesian study employed the MARSI questionnaire to investigate the types and frequency of metacognitive reading strategies used by both high and low-performing students (Aziz et al., 2019). The questionnaire was distributed to students, and their responses were analyzed to identify the strategies used and their frequency. The study revealed that high-performing students used a broader array of metacognitive reading strategies and employed them more frequently compared to their low-performing peers (Aziz et al., 2019). However, this study was limited by its descriptive statistics and qualitative approach, coupled with a small sample size that failed to provide a representative group of participants. An additional concern pertains to the use of an ordinal scale for mean calculation without applying the necessary measurement process, as demonstrated in the Rasch measurement. Therefore, given these limitations, further research using appropriate measurement to examine individual awareness and use of metacognitive reading strategies in the education field is crucial to provide an excellent example for future researchers.

1.3. Rasch analysis and rating scale improvement

In Rasch analysis, the impact of ability on the difficulty of the item is measured, "making the item and person estimates inter-dependent" (Oon & Fan, 2017, p. 4). Rasch analysis is useful for characterizing a test's rating scale and the test-takers ability, as well as identifying the quality of the psychometric features for the capacity of the evaluation scale (Oon & Fan, 2017; Rasch, 1993). The rating scale in Rasch analysis examines the difficulty of the question item and the progress made or the ability level of the performer or learner in a certain field (see e.g., Oon & Fan, 2017). High values show higher difficulty, and lower ones show lower difficulty (Rasch, 1993).

According to Pesudovs and Noble (2005), Rasch analysis "provides a method for testing scale assumptions and modifying scale structure to become a truly linear scale" (p. 631), enabling further improvement to the instruments, as the newly made ones would no longer require additional testing or examination (Pesudovs & Noble, 2005). Other researchers (Koopmans et al., 2014) have claimed that Rasch analysis "provides information on whether a questionnaire can measure persons at all levels of ability, in the form of person-item distribution maps" (p. 161). Likewise, the scale of the person-item can be represented in a map "from −5 to + 5, with 0 being the average item difficulty" (Koopmans et al., 2014, p. 163), which also indicates how suitable the items are for the person. Improvements to the rating scale are generally made by removing items with "a high difficulty parameter" and with a "low difficulty parameter," after which the improved scale is "re-estimated accordingly"; the ultimate scale of the items should fall in a range

between < 2.5 and > −2.5 (Koopmans et al., 2014, p. 163).

Another reason why Rasch analysis is used in this study is to address certain drawbacks associated with Classical Test Theory (CTT). There are four main limitations of CTT when defining a measurement model: (1) measurements are formulated using ordinal data results rather than an interval scale (logit); (2) both the item and person in the measurement process are interdependent; (3) the properties of the measurement tool, specifically reliability and validity, greatly depend on the sample; and (4) the data is focused on group-centered statistics, which are not suitable for illustrating the measurement of individual respondents (Barbic & Cano, 2016).

Moreover, Rasch analysis can be used to assess a person's response to items in a developed or adapted instrument as long as the data are dichotomous or polytomous data (Andrich, 2018). Several psychometric properties can be examined to determine the validity and reliability of the instruments used. Person and item fit statistics, infit and outfit mean square (MNSQ) are used to confirm whether the data match the expectations of the Rasch model. The item characteristic curve (ICC) indicates how well an item discriminates between people of different ability levels. The steeper the curve, the better the item discriminates between different ability levels. Differential Item Functioning (DIF) is used to check whether different groups respond differently to an item,
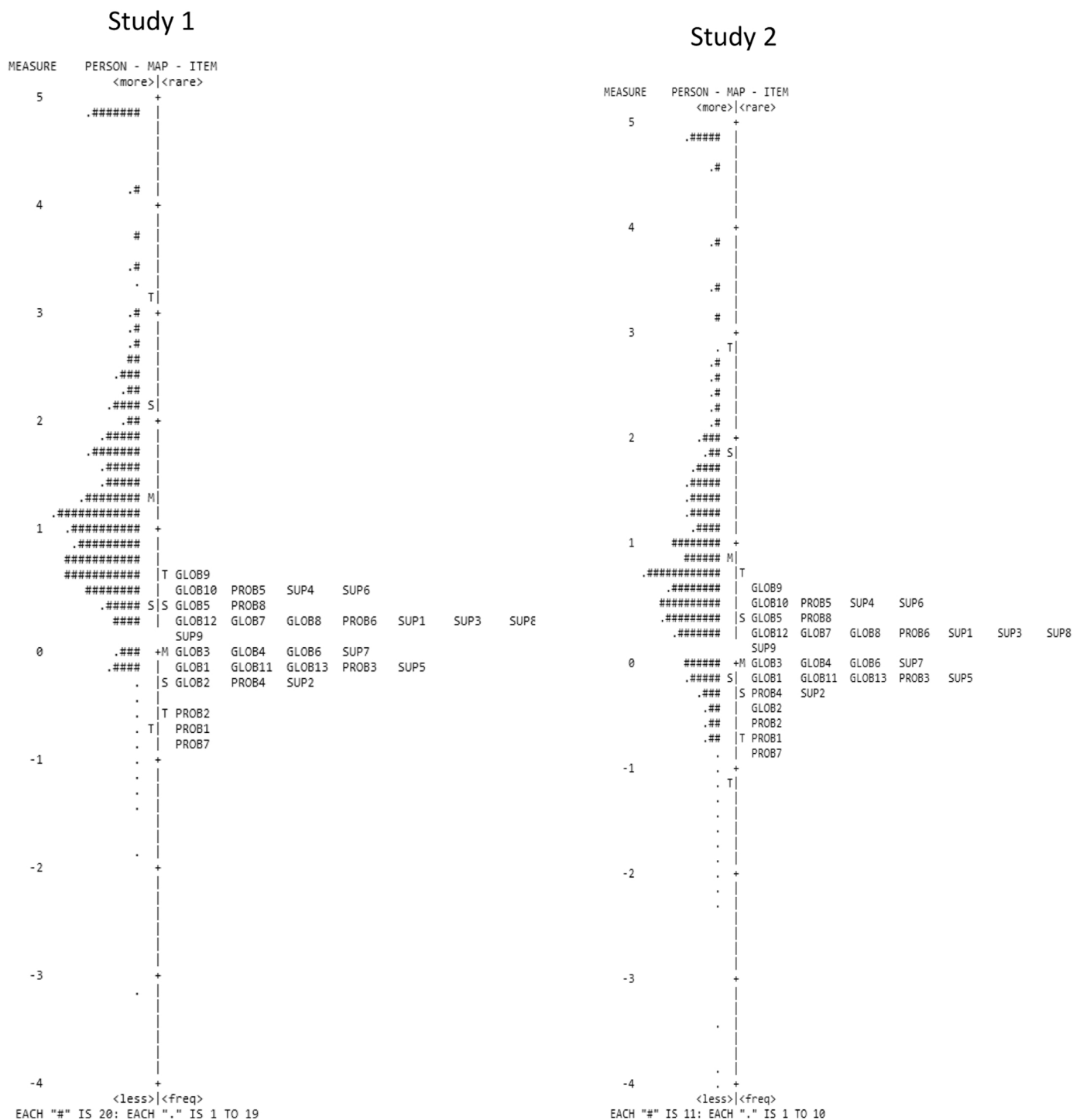

Fig. 1. Item-person maps. Note: (#),20 participants, (.) 1–19 participants.

confirming any indication of item-level bias. Unidimensionality checks the fit of the Rasch model by assessing the raw variance explained by the measure to confirm that the residuals should be random and not show any pattern. Local independence looks at the correlation between residuals. If items are locally independent, the residuals should not be correlated (Bond et al., 2020; Boone et al., 2014; Linacre, 2021).

The discourse on the optimal number of rating points on a Likert scale has emerged as a significant concern within the field of educational psychology (Cowan, 2015; Hair et al., 2020; Miller, 1956; Pesämaa et al., 2021; Saaty & Ozdemir, 2003). Dawes (2008) has concluded that ten-point scales do not alter variability, while other methodological scholars argue that 11-point scales (0−10) do indeed increase variability (Hair et al., 2020; Pesämaa et al., 2021). However, Miller (1956) noted a limit to the human capacity for information processing, suggesting that a 7-point Likert scale represents the maximum amount of information one can handle. This finding is reinforced by Saaty (2003), who recommended that questionnaires with either 5 or 4 rating points are most comprehensible due to their widespread use. Nonetheless, these previous studies could not definitively determine the optimal number of scales for a questionnaire, as they merely compared variability and consistency using a group-centered statistical approach. In this study, we endeavor to ascertain the number of rating points that respondents can accurately comprehend using the MARSI Questionnaire, via an individual-centered statistical approach incorporating Rasch measurement, following the guidelines by Andrich (2018), Andrich and Marais (2019) and Linacre (2021). Findings from the rating scale analysis in this study indicate that respondents more easily understood the MARSI questionnaire when it featured a 4-point rating scale, based on the probability curve in Fig. 4 and the rating scale function in Table 4. Both Fig. 4 and Table 4 confirm that there are no overlapping lines for each rating point and that the Andrich threshold values increase monotonically. Therefore, we hope the rating scale analysis using the Rasch measurement application in this study will encourage future research in educational contexts to ensure the robust validity of evidence, specifically concerning individuals' understanding of the number of rating points.

1.4. Research questions (RQ)

This study evaluated the psychometric properties of the Indonesian version of the Metacognitive Awareness of Reading Strategies Inventory (MARSI) questionnaire from pre-service primary teachers and investigated pre-service primary teachers' metacognitive awareness of reading strategies. The data from Study 1 and Study 2 were analyzed using the Rasch measurement approach via WINSTEPS 5.2.5.1 software (Linacre, 2022). The following four research questions were created to guide study purposes:

RQ1. Does the Indonesian translation of the MARSI questionnaire meet reliability and validity criteria based on Rasch modeling?

RQ2. Is there any significant item bias detected based on gender using DIF analysis?

RQ3. How is the rating scale category functioning based on a comparison of Study 1 with five rating scales and Study 2 with four category scales?

RQ4. How do pre-service primary teachers' abilities categorize based on metacognitive awareness of reading strategies?

2. Methods

2.1. Participants and procedures

A quantitative method and a cross-sectional research design were utilized in this study, adopting a two-step data collection procedure. In study 1, 2912 Indonesian students participated voluntarily by completing the Indonesian version of the MARSI questionnaire, using with five rating scale categories, following the original English version,

ranging from 1 (never) to 5 (always). In study 2, 1465 Indonesian student participants were also filled out voluntarily after we altered the scale from 5 to 4 points using the Indonesian version of the MARSI questionnaire. The data were collected via an online questionnaire tool, and the respondents represented four major islands of Indonesia, namely Kalimantan, Java, Sumatera, and Sulawesi. In Indonesia, there are two primary paths to major in primary teacher education. The first involves the standard admission process into the teacher training and education department with the primary education major. The second is participation in government initiatives specifically designed for existing primary school teachers who do not yet possess a bachelor's degree in teacher training and education. We are specifically seeking study participants who are aged between 21 and 40 years. As such, it is expected that all participants have completed at least two semesters at the university level. All participants who meet these criteria will be included in our dataset without any additional exclusion criteria. Written consent from participants was obtained before they completed the questionnaire, and their responses were treated as confidential and anonymous. Participants were reached through various channels, including university emails, WhatsApp groups, and their lecturers. They were asked to complete an online form of the MARSI Questionnaire, which is available in the Indonesian language. It is essential that all participants fill out each question in the questionnaire without skipping any items. All measured variables are discussed in the present study as presented in Table 1. The university type and living places were collected, but they will be analyzed and discussed in a different publication. In this study, we only focus on gender and age investigation.

2.2. Instruments

Permission to use the MARSI Questionnaire was granted by Kouider Mokhtari, Ph.D, the copyright holder of the questionnaire via personal communication. This consent allows for the questionnaire to be used strictly for the academic purposes of this study. This is the first validation of the MARSI Questionnaire in the Indonesian version with the author's permission as a copyright holder that applies Rasch measurement and performs rating scale improvement analysis for empirical study. Background Information was collected on the participants' gender, living place, university type, and age. The MARSI questionnaire (Mokhtari & Reichard, 2002) was used to assess participants' reading awareness and their perception of the reading strategies they use while reading textbooks or particular subject books in their school curricula. The MARSI consists of 30 items, with responses given on a 5-point Likert scale from 1 (never) to 5 (always). The MARSI has three dimensions: Support Reading Strategies (SUP) (9 items), Global Reading Strategies (GLOB) (13 items), and Problem-Solving Strategies (PROB) (8 items). SUP represents individual reading strategies for capturing critical information, summarizing information, and taking notes related to specific information found in academic materials. GLOB focuses on the purposes of reading, previewing the text, using tables and figures, skimming, and deciding the essential information. PROB is related to how the reader adjusts reading speed, guesses the meanings of unfamiliar words, and visualizes concepts. All instruments were provided to the participants in their Indonesian versions. Prior to test administration, we checked the content validity of the instruments through back-and-forward translation, which was carried out by two Indonesian native speakers who specialize in the English language.

2.3. Data analysis

SPSS version 25 (IBM Corp, 2017) was employed to produce descriptive statistics for participant demographics. WINSTEPS version 5.2.5.1 software (Linacre, 2022) was used for Rasch analysis with joint maximum likelihood estimation (JMLE) estimation to perform psychometric evaluations and ratings. Rasch modeling was used in this study, and the scores of the students were transformed into interval data
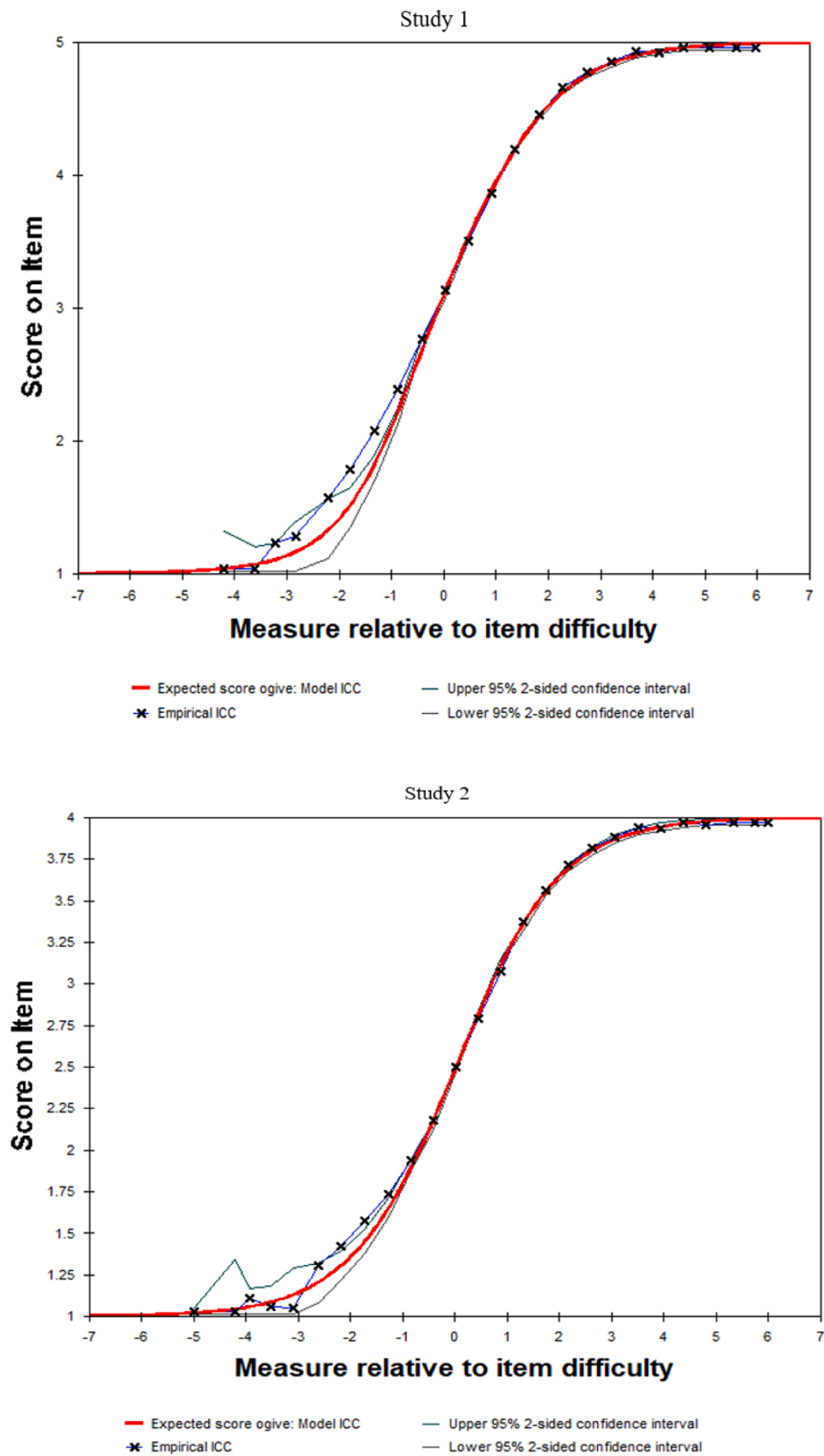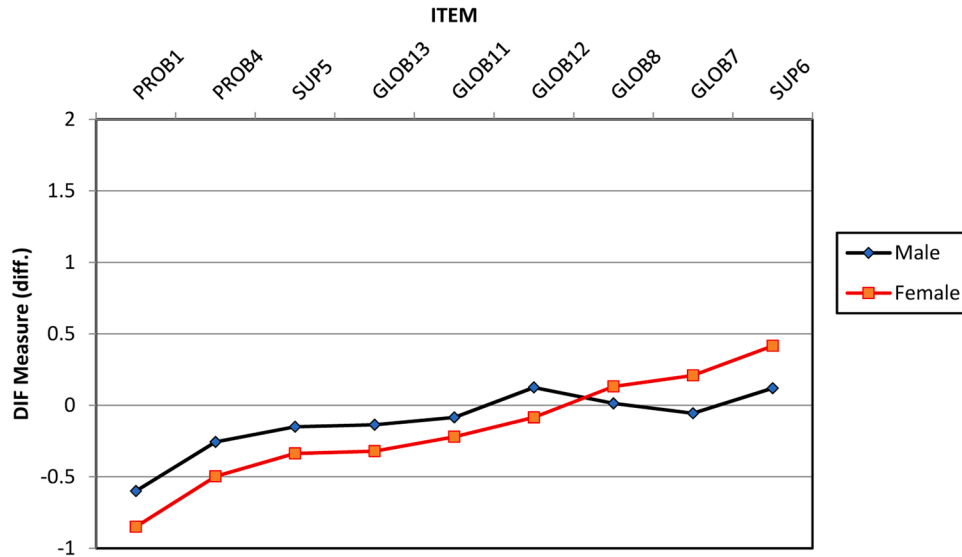
Fig. 2. ICC plot for study 1 and study 2.

DIF analysis based on gender for study 1

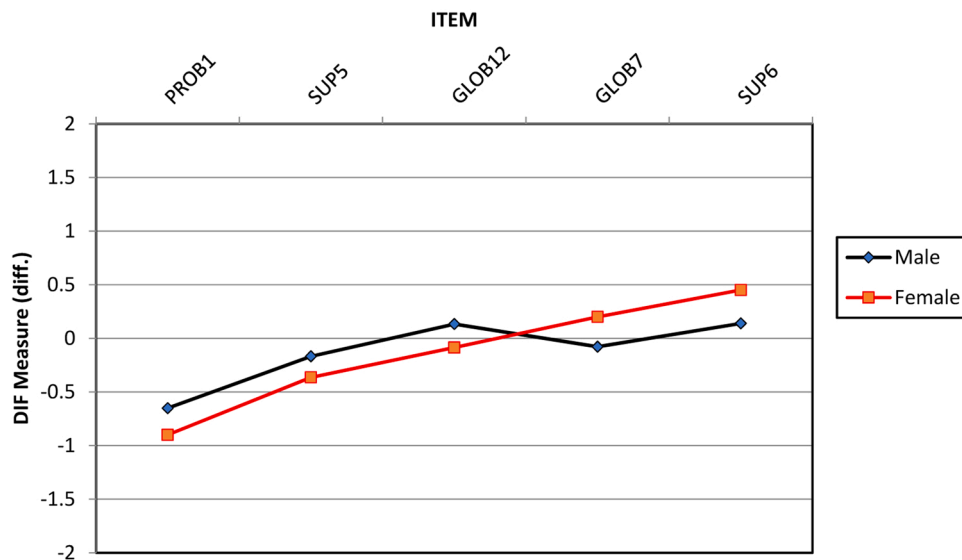## DIF analysis based on gender for study 2

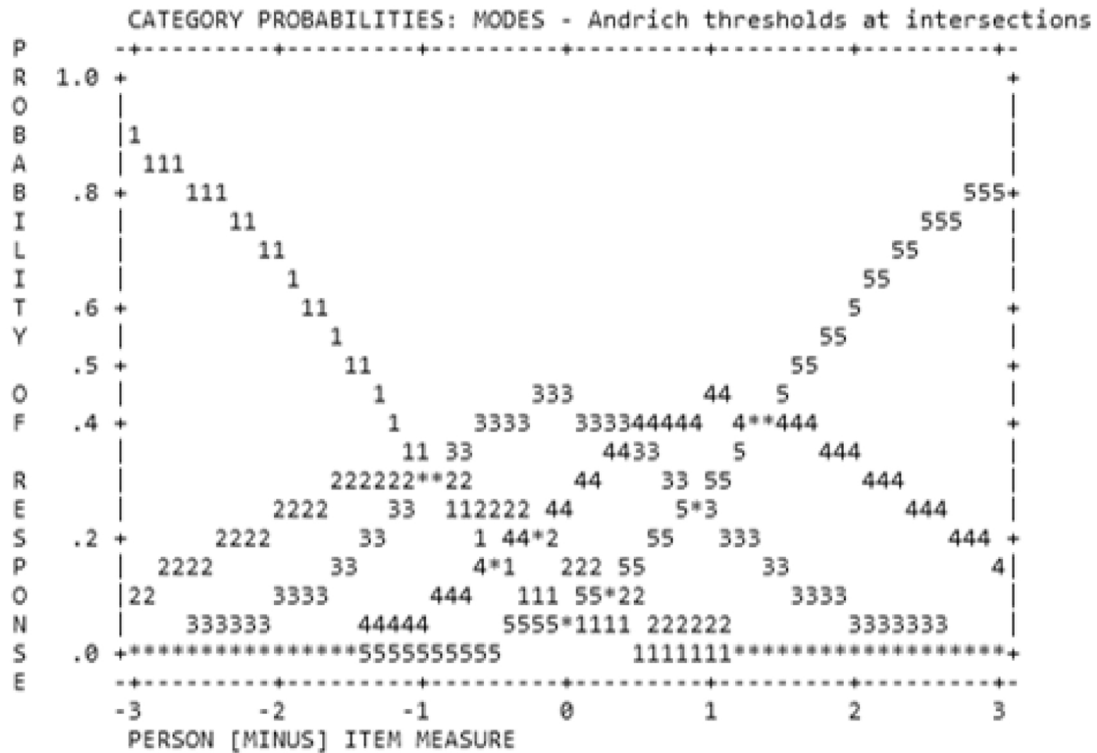**Fig. 3.** DIF analysis based on gender for Study 1 and Study 2.

(logits) ranging from negative infinity to positive infinity.

Several Rasch indicators based on the unidimensionality, the local independence, item and person separation, item reliability, and the fit validity based on infit and outfit mean MNSQ values were used to investigate measurement properties of MARSI questionnaire. The unidimensionality can be confirmed by checking the raw variance values by measure for all tasks. The raw variance values by measures should be above 30% and unexplained variance first contrast below 2 (Linacre, 2021). The raw residual correlation based on Yen's Q3 statistic (Christensen et al., 2017) was used to confirm no local independence was found, with a raw residual correlation below 0.4. The values for item and person separation should exceed 2 logits. This requirement ensures that there are at least two distinct group levels present within the dataset (Bond et al., 2020). The item reliability and Cronbach's alpha (α) values should be above 0.6 to ensure reliability hold in this study (Fisher, 2007; Taber, 2018). The fit validity can be evaluated using infit and outfit

mean MNSQ values whereby the acceptable range for these indicators is from 0.5 to 1.5, although 1.6 is still regarded as an acceptable criterion if the point measure correlation (PTMA) is still positive (Boone et al., 2014; Park & Liu, 2019).

An item-person map was developed to confirm the interaction between items and persons. The ICC plots were determined by group to ensure the fit validity of the items at the instrument level. DIF analysis was applied on gender to verify the bias interaction for each item. We also compared the rating scale functioning between study 1, with a 5-point scale, and study 2, with a 4-point scale, to determine relative performance for the scales in an Indonesian context. Finally, to investigate respondents' metacognitive awareness of their reading strategies, respondent abilities were compared using a histogram with logit measures using R statistics. The logit value of a person (LVP) was investigated using Microsoft Excel with the COUNTIF formula to classify mean logits and standard deviations.

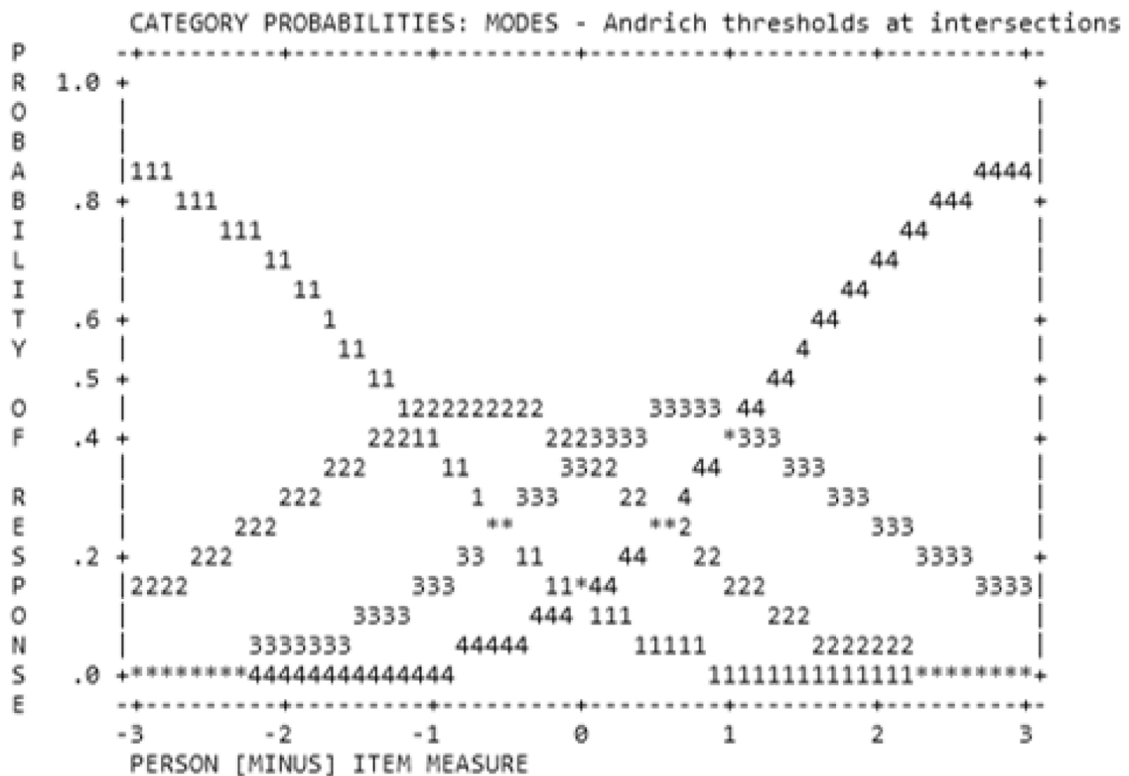**Fig. 4.** Probability curve of MARSI scale for study 1 and study 2.

Table 1
Demographic profiles of pre-service primary teachers in this study.

Demographic		Study 1		Study 2	
		Frequency	Percentage (%)	Frequency	Percentage (%)
Gender	Male	1189	40.8	653	44.6
	Female	1723	59.2	812	55.4
Age	below 25 years old	1848	63.5	861	58.8
	25–30 years old	761	26.1	358	24.4
	31–35 years old	259	8.9	122	8.3
	above 35 years old	44	1.5	124	8.5
University type	Public	1953	67.1	981	67.0
	Private	959	32.9	484	33.0
Living place	City	742	25.5	375	25.6
	District	2170	74.5	1090	74.4

3. Result

3.1. Validity and reliability of the MARSI questionnaire (RQ1)

3.1.1. Validity

To confirm the validity of the MARSI questionnaire, the item and person parameters produced using Rasch analysis were assessed with the infit and outfit MNSQ as presented in Table 2. The infit and outfit MNSQ values indicate items and persons achieving fit validity criteria. Because this study had a large sample size of more than 500 students, the infit and outfit z-standardized (ZSTD) can be ignored as a threshold for fit validity criteria (Azizan et al., 2020). The separation values for item and person should be greater than 2 logits to confirm that there are more than two different groups in terms of individual ability and item difficulty level.

Table 2 shows the summary of Rasch parameters in study 1 and study 2 for the MARSI questionnaire in the Indonesian version, indicating that the person and item level achieve fit validity criteria in both studies, based on MNSQ values ranging from 1.00 to 1.06. Item and person separations also feature values above 2 logits. The construct validity of the MARSI was confirmed by assessing instrument unidimensionality and local independence. The result of raw variance values in Table 2 confirms that Unidimensionality was attained in both studies.

Table 3 shows the item measures and fit criteria to ensure fit validity

at the item level. The item measures in Study 1 ranged from 0.80 logits to 0.51 logits, and the Outfit MNSQ values ranged from 0.76 logits to 1.68 logits. Item PROB6, with 1.68 logits for Outfit MNSQ values, is retained due to its positive PTMA. Study 2 shows a similar result, with item measures ranging from -0.85 logits to 0.55 logits and Outfit MNSQ values ranging from 0.77 logits to 1.42 logits. Thus, the Indonesian version of the MARSI questionnaire is valid for all items in both studies.

An item-person map is used to investigate the interaction between items and persons in both studies. As shown in Fig. 1, the majority of respondents have greater ability above the item difficulty level, which confirms that pre-service primary teachers had a good metacognitive awareness of reading strategies. We also develop ICC plots based on the level of the instrument. The ICC plot confirms that both studies fit the Rasch probability whereby the empirical and expected lines match or overlap with each other.

3.1.2. Reliability

Item reliability criteria were determined using Cronbach's alpha (α) for whole items in the MARSI questionnaire and for each dimension (see Table 2). The item reliability values ranged from 0.98 to 0.99, confirming reliability on the item level for both studies (Fisher, 2007). Cronbach's alpha (α) values ranged from 0.82 to 0.94, with 0.6 representing the minimum threshold (Taber, 2018). Based on Cronbach's alpha values, these results indicate that reliability was also attained.

3.2. DIF analysis based on gender (RQ2)

DIF analysis was used to ensure whether items displayed gender bias (female and male) that would affect pre-service primary teachers' reading strategy abilities. DIF analysis can indeed distinguish participant bias at the item level in the questionnaire based on subgroups or background variables (Boone et al., 2014; Khine, 2020). DIF result was calculated using two criteria: significant probability ($p < 0.05$) and DIF contrast. There are three DIF contrast classifications used (Zwick et al., 1999), namely, negligible, slight to moderate ($| \text{DIF} | \geq 0.43$ logits), and moderate to large ($| \text{DIF} | \geq 0.64$ logits). Fig. 3 depicts the DIF size based on significant probability for several items in both studies. However, no items have a DIF contrast greater than 0.43 logits. As a result, we can conclude that all items have negligible DIF, indicating that the instrument has no bias issues based on gender.

3.3. Rating scale category functioning between study 1 and study 2 (RQ3)

The rating scale functioning for the 5-point scale for the Indonesian version of the MARSI questionnaire in study 1 was compared to the 4-

Table 2
Summary of Rasch parameters for MARSI in Study 1 and Study 2.

Psychometrics Attribute	Study 1				Study 2			
	Subscale			MARSI	Subscale			MARSI
	GLOB	PROB	SUP		GLOB	PROB	SUP	
Number of Items	13	8	9	30	13	8	9	30
Mean								
item outfit MNSQ	1.03	1.04	1.01	1.05	1.02	1.01	1.01	1.04
item Infit MNSQ	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
person outfit MNSQ	1.03	1.01	1.01	1.05	1.02	1.01	1.01	1.04
person Infit MNSQ	1.03	1.01	1.01	1.06	1.02	1.03	1.02	1.05
Item separation	9.89	17.81	9.60	13.12	7.19	12.97	6.85	9.48
Person separation	2.09	2.67	1.93	3.36	2.18	1.96	2.72	3.41
Item Reliability	.99	1.00	.99	.99	.98	.99	.98	.99
Cronbach's Alpha	.86	.82	.85	.93	.86	.83	.86	.94
Unidimensionality								
Raw variance explained by measure	34.1%	41.7%	40.2%		33.6%	41.4%	39.9%	
Unexplained variance 1st contrast	1.72	1.49	1.78		1.68	1.47	1.78	

Table 3
Item measures and fit criteria.

Item number	Study 1			Study 2		
	measure (logits)	Outfit MNSQ	PTMA	measure (logits)	Outfit MNSQ	PTMA
GLOB1	− 0.14	1.13	0.49	− 0.14	1.11	0.52
GLOB2	− 0.35	0.83	0.51	− 0.38	0.84	0.53
GLOB3	0.03	1.10	0.50	0.03	1.09	0.52
GLOB4	0.03	1.13	0.51	0.04	1.12	0.53
GLOB5	0.25	1.03	0.52	0.28	1.03	0.54
GLOB6	− 0.04	1.68	0.41	− 0.07	1.61	0.42
GLOB7	0.19	1.26	0.49	0.18	1.25	0.51
GLOB8	0.17	0.92	0.54	0.18	0.91	0.56
GLOB9	0.51	1.42	0.48	0.55	1.40	0.49
GLOB10	0.40	0.81	0.58	0.46	0.86	0.60
GLOB11	− 0.08	0.77	0.56	− 0.09	0.77	0.58
GLOB12	0.09	1.04	0.52	0.1	0.99	0.55
GLOB13	− 0.16	1.50	0.43	− 0.19	1.42	0.46
PROB1	− 0.67	1.07	0.46	− 0.71	1.06	0.49
PROB2	− 0.58	0.93	0.48	− 0.62	0.90	0.51
PROB3	− 0.18	1.15	0.48	− 0.21	1.12	0.50
PROB4	− 0.31	0.96	0.50	− 0.35	0.93	0.52
PROB5	0.38	1.18	0.52	0.41	1.17	0.53
PROB6	0.11	0.89	0.55	0.12	0.89	0.57
PROB7	− 0.80	0.84	0.48	-0.85	0.84	0.50
PROB8	0.23	1.27	0.49	0.26	1.25	0.51
SUP1	0.20	1.06	0.53	0.21	1.06	0.55
SUP2	− 0.22	0.76	0.55	-0.24	0.75	0.58
SUP3	0.11	0.97	0.54	0.13	1.01	0.55
SUP4	0.36	1.10	0.54	0.39	1.06	0.56
SUP5	-0.18	1.01	0.50	− 0.19	1.02	0.52
SUP6	0.39	1.04	0.53	0.43	1.08	0.55
SUP7	-0.04	0.76	0.57	− 0.04	0.77	0.59
SUP8	0.11	0.81	0.57	0.14	0.84	0.59
SUP9	0.18	1.08	0.51	0.21	1.10	0.53

point scale in study 2 to determine how well respondents understood the scales. The category probability curve in Fig. 4 indicates that the 4-point scale in Study 2 had a clearer peak than the 5-point scale in Study 1. Thus, the Indonesian version of the MARSI questionnaire with a 4-point rating scale was more easily understood by respondents (Fisher, 2007). Further, as shown in Table 4, study 2 had no disordered thresholds (steps) whereby the observed average measure by category improved monotonically, confirming that all scale categories were well functioning (Andrich, 2018; Andrich & Marais, 2019; Fisher, 2007).

3.4. Evaluation of pre-service primary teachers' metacognitive awareness of reading strategies (RQ4)

R statistics integrated into WINSTEPS software was used to determine pre-service primary teachers' ability and determine the histogram for both studies (Fig. 5). In study 1, individual abilities ranged from − 4 logits to 7 logits, and the majority of respondents (2125) had logit values from 0 logits to 2 logits. Study 2 shows a similar result, where individual

abilities ranged from − 6 logits to 6 logits, and the majority of respondents are above 0 logits. This result relates to the item-person map, whereby most pre-service primary teachers had good ability above average values (0 logits 9 in the metacognitive awareness of reading strategies.

LVP analysis was used to evaluate the categorization of pre-service primary teachers' ability to complete the MARSI questionnaire adequately. In study 1, the following were calculated: the mean logit of the person (1.42 logits), 1 SD (1.35 logits), − 1 SD (−1.35 logits), mean logit + 2 SD (4.12 logits), and the mean logit + 2 SD (−1.28 logits). In study 2, the following were calculated: the mean logit of the person (0.87 logits), 1 SD (1.03 logits), − 1 SD (−1.03 logits), mean logit + 2 SD (2.93 logits) and the mean logit + 2 SD (−1.19 logits). From the mean logit of the person and SD values, pre-service primary teacher abilities were categorized into very high, high, moderate, and low. Table 5 summarizes the LVP analysis performed based on gender and age, using the COUNTIF function in Microsoft Excel to operate an automatic estimation of participant logit measures.

According to the findings in Table 5, in Study 1, 108 (3.7%) females and 72 (2.47%) males were classified as possessing very high abilities. Moreover, 571 (19.60%) females and 285 (9.78%) males were identified as having high abilities. Furthermore, 1042 (35.78%) females and 826 (28.36%) males were categorized as having moderate abilities, while 2 (0.06%) females and 6 (0.21%) males were classified as having low abilities. In study 2, teacher abilities were classified as follows: 40 (2.73%) females and 33 (2.25%) males with very high abilities, 200 (13.65%) females and 129 (8.80%) males with high abilities, 562 (38.22%) females and 483 (32.96%) males with moderate abilities, and 10 (0.68%) females and 8 (0.54%) males with low abilities.

Demographic characteristics were collected. In study 1, 110 (3.7%) pre-service primary teachers were below 25 years in a total of 180 respondents classified into the high category, and 536 (18.40%) preservice primary teachers were below 25 years in a total of 856 respondents had a high category as well. Almost half of the respondents (41.07%) with ages below 25 years were classified into the moderate category. Only 6 (0.2%) pre-service primary teachers with ages below 25 years and 2 (0.06%) pre-service primary teachers with ages 31–35 years fall into the low category. In study 2, the majority of pre-service primary teachers with ages below 25 years were classified into the moderate category (41.91%). The details of this categorization can be seen in Table 5.

4. Discussion

This study established validity and reliability based on Rasch measurement for the Indonesian version of the MARSI questionnaire in two studies. In addition, instrument enhancement was also investigated in terms of performance rating (Table 4 and Fig. 4). It was found that a 4-point scale for the MARSI Questionnaire was more suitable for the preservice primary teachers, the respondents in this study. Rating
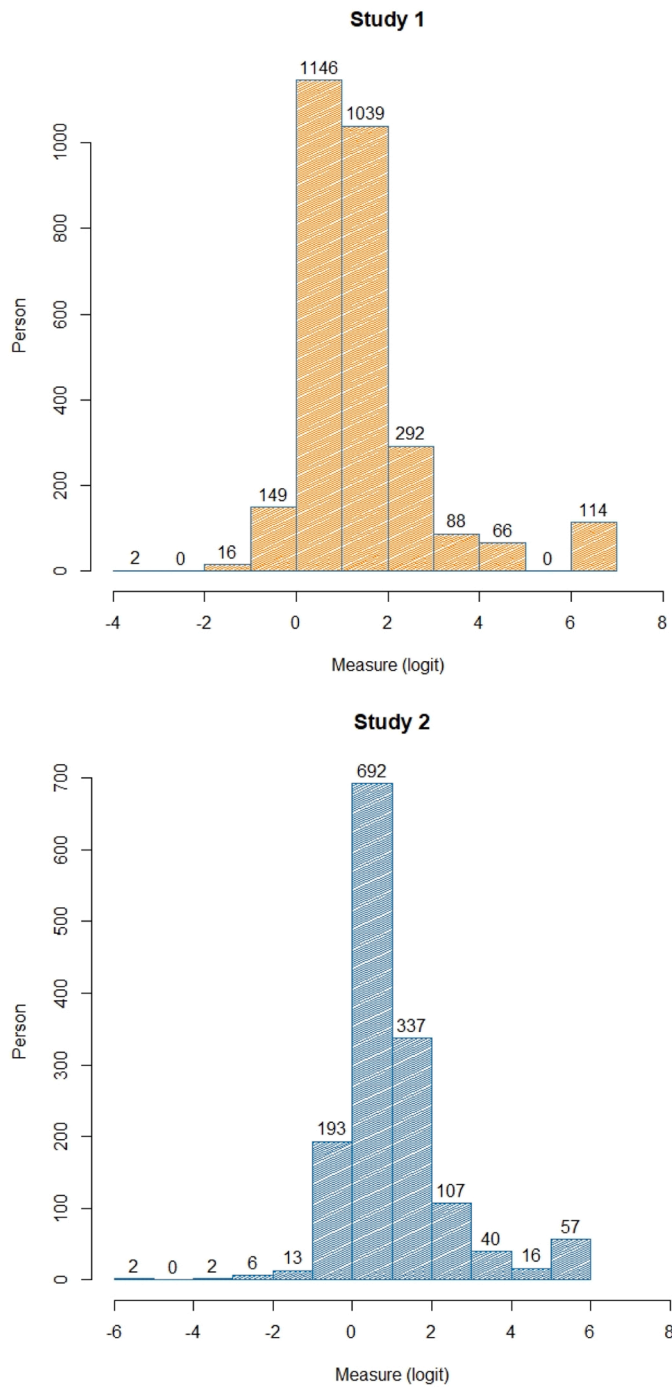
Table 4
Rating scale functioning.

Category label	Observed		Observed average	INFIT MNSQ	OUTFIT MNSQ	Andrich Threshold
	Count	%				
Study 1						
1 (Never)	2279	2	0.14	1.39	1.83	NONE
2 (Only occasionally)	5801	6	0.22	1.03	1.11	− 0.93
3 (Sometimes)	22285	23	0.55	0.89	0.91	− 0.92
4 (Usually)	32684	33	1.04	0.85	0.82	0.45
5 (Always)	35959	36	1.89	0.98	0.98	1.4
Study 2						
1 (Never)	4054	8	-0.21	1.21	1.37	NONE
2 (Only occasionally)	11221	23	0.11	0.91	0.95	− 1.12
3 (Usually)	16458	33	0.67	0.88	0.83	0.05
4 (Always)	18077	36	1.58	0.97	0.99	1.07

Fig. 5. Pre-service primary teachers' ability on metacognitive awareness of reading strategies for study 1 and study 2.

Table 5
LVP analysis for pre-service primary teachers' metacognitive awareness of reading strategies.

Demographics	Very high, LVP > Mean Logit + 2 SD	High, Mean Logit + 2 SD ≥ LVP ≥ Mean Logit	Moderate, Mean Logit ≤ LVP ≤ Mean Logit - 2 SD	Low, LVP < Mean Logit - 2 SD
Study 1				
Gender				
Male	72	285	826	6
Female	108	571	1042	2
Total	180	856	1868	8
Age				
below 25 years old	110	536	1196	6
25–30 years old	38	201	522	0
31–35 years old	28	97	132	2
above 35 years old	4	22	18	0
Total	180	856	1868	8
Study 2				
Gender				
Male	33	129	483	8
Female	40	200	562	10
Total	73	329	1045	18
Age				
below 25 years old	43	192	614	12
25–30 years old	14	69	273	2
31–35 years old	11	35	75	1
above 35 years old	5	33	83	3
Total	73	329	1045	18

questionnaire attaining consistency in measurement. In addition, the results of this study prove that The MARSI questionnaire can be used properly, covering the weaknesses of previous studies in the Indonesian context (Aziz et al., 2019; Deliany & Cahyono, 2020; Halim et al., 2022) where the assessment process was not accompanied by robust psychometric evidence and large sample size. Fig. 3 indicates the result of bias issues based on gender. DIF analysis is used to investigate bias in the instrument. In this study, we found that the instruments in study 1 and study 2 are free of bias issues using DIF analysis, as shown in Fig. 3. This finding contributes to explaining the area of bias measurement that was uncovered in previous research by Mokhtari & Reichard (2002). Previous studies from Dinçer & Çilek (2022) provided a comparison between female and male abilities. However, this result does not clarify issues of bias because it focused only on summing the raw data and using means comparison. Rezat et al. (2022) also investigated pre-service primary teachers in mathematics education, but no DIF analysis was performed to check bias issues in that context. Therefore, we hope that the results of this study can help other researchers to assess metacognitive awareness of reading strategies or validate instruments using Rasch measurement.

The reading strategy of pre-service primary teachers needs to be assessed because primary education plays a vital role in fostering students' interest in reading and learning (Rezat et al., 2022; Skaar et al., 2018). The evaluation of pre-service primary teachers' metacognitive awareness of reading strategies using LVP analysis and histograms with R statistics integration, as presented in Fig. 5 and Table 5. The results confirm that the majority of pre-service primary teachers place an above-average level in reading strategies based on Rasch scaling. This result is in line with Dinçer & Çilek (2022), who evaluated metacognitive awareness of reading strategies and critical thinking in pre-service classroom teachers and discovered that the mean metacognitive awareness of reading strategies of pre-service teachers was higher than 3 out of 5 on the Likert scale. Iwai (2016) also provided a

improvement analysis is rarely performed in the educational context. We followed precedent in health science applications (Axon et al., 2019) and employed Rasch analysis. In some recent studies (Deliany & Cahyono, 2020a; Halim et al., 2022; Kallio et al., 2018; Nahar & Mallik, 2022; Shah & Asgher, 2022; Upadhyay et al., 2022; Villanueva, 2022), the MARSI questionnaire used in educational context has not provided comprehensive and complex evidence to corroborate the validity and reliability of the questionnaire, primarily in terms of validity checks at the item level and avoiding bias issues. Table 3 shows the area of validity at the item level by comparing two studies with different data collections. The results of item fit measures are valid based on the Rasch measurement and not significantly different, which confirms the MARSI

similar finding from 116 pre-service teachers, in which the average reading strategies scores were above average, 3.35 out of 5. Intriguingly, Dinçer & Çilek (2022) and Iwai (2016) analyzed scores on a Likert scale, using raw scores and summing the scores to get an average score. This practical analysis has become a subject of debate among researchers, as Likert scales are ordinal scales in which summing and averaging are not appropriate if Rasch measurement has not been undertaken (Jamieson, 2004; Leung, 2011; Pimentel & Pimentel, 2019; Wu & Leung, 2017). Rasch measurement converts the ordinal scales into interval scales called logits to pass the barrier of ordinal scales. Therefore, it is possible to check the results in all indicators in item and instrument levels.

To categorize pre-service primary teachers' abilities in reading strategy, LVP analysis was performed using the recommendations from Soeharto and Csapó (2022) and Chan et al. (Chan et al., 2021). For this study, pre-service teachers' abilities in terms of reading strategy for study 1 and study 2. A significant proportion of pre-service primary teachers are classified as having moderate abilities, whereby females are greater than males. This finding supports previous studies that have found females tend to have better reading strategies than males (Deliany & Cahyono, 2020; Dinçer & Çilek, 2022; Jaleel & P., 2016; Kallio et al., 2018; Shah & Asgher, 2022; Upadhyay et al., 2022). In study 1, almost half of the respondents in study 1 and study 2 fall into the moderate category. 41.07% of pre-service primary teachers with ages below 25 years old were classified into the moderate category in Study 1. Also, 41.91% of the majority of pre-service primary teachers with ages below 25 years old were classified into the moderate category. Surprisingly, based on the evaluation in Table 5 and Fig. 5, we can assume that pre-service primary teachers in Indonesia have good reading strategies. This result corroborated previous studies in different contexts that found pre-service teacher reading strategies at the university level had high and moderate categories in metacognitive awareness of reading strategy (Halim et al., 2022; Villanueva, 2022).

5. Limitations and recommendations for future study

This study had third main limitations. First, the participants were only drawn randomly from four major islands in Indonesia, namely, Kalimantan, Java, Sumatera, and Sulawesi, where suitable Internet connections could be found to facilitate data collection. However, we have used random sampling to increase the representativeness of the data collected. Future studies should seek to collect data using a paper-based test with collaboration from a teacher's group association to provide a complex and comprehensive dataset. Second, no treatment or experiment in this study was performed to make a comparison study. This study only applied a cross-sectional study using a quantitative method, assisted by Rasch measurement. Therefore, future studies should attempt to apply different training programs related to reading and compare the results of respondents' metacognitive awareness of reading strategies. Finally, we recommend other researchers investigate the relationship between metacognitive awareness of reading strategies and academic achievements in different subject areas, as recommended by Young and Fry (2008). Third, it is possible that language and cultural differences may affect the MARSI questionnaire, which was originally developed in English. However, we have minimized this effect by reviewing the content and employing two language experts who used back-and-forward translation. Regarding Rasch measurements, there is a possibility of checking for language bias by applying the questionnaire in two different languages. Therefore, conducting further studies related to language bias in adapted questionnaires is recommended.

6. Conclusions

This study confirmed that the Indonesian version of the MARSI questionnaire is valid and reliable based on Rasch parameters on the instrument and the item levels. Item-person maps revealed that most respondents had a good metacognitive awareness of reading strategies

or above-average levels (0 logits). A DIF analysis showed that no substantial DIF size was detected for any items, confirming no bias issue based on gender. The MARSI questionnaire using a 4-point Likert scale was better understood by respondents than the version of the MARSI questionnaire with a 5-point scale. The category probability curve showed that for the MARSI questionnaire for the 4-point scale with no disordered thresholds (steps), the observed average measure by category improved monotonically. LVP analysis illustrated pre-service primary teachers' abilities in reading strategy for both study 1 and study 2. Most pre-service primary teachers fall into the moderate category regarding metacognitive awareness of reading strategy ability.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgement

References

Akhmetova, A., Imambayeva, G., & Csapó, B. (2022). Reading strategies and reading achievement in middle school: Kazakhstani Young Learners. *SAGE Open, 12*(3), 1–15. https://doi.org/10.1177/21582440221113843

Akyol, H., & Ulusoy, M. (2010). Pre-service teachers' use of reading strategies in their own readings and future classrooms. *Teaching and Teacher Education, 26*(4), 878–884. https://doi.org/10.1016/j.tate.2009.10.026

Aziz, Z. A., Nasir, C., & Ramazani, R. (2019). Applying metacognitive strategies in comprehending english reading texts. *Celt: A Journal of Culture, English Language Teaching & Literature, 19(1), Article*, 1. https://doi.org/10.24167/celt.v19i1.1863

Andrich, D. (2018). Advances in social measurement: A Rasch measurement theory. *In Perceived Health and Adaptation in Chronic Disease* (pp. 66–91). Routledge,.

Andrich, D., & Marais, I. (2019). A Course in Rasch Measurement Theory: Measuring in the Educational. *Social and Health Sciences.* Springer Nature Singapore Pte Ltd.,.

Axon, D. R., Augustine, J. M., Warholak, T., & Lee, J. K. (2019). Improving rating scales: Applying Rasch analysis to student pharmacists' attitudes towards herbal medications. *Currents in Pharmacy Teaching and Learning, 11*(7), 658–663. https://doi.org/10.1016/j.cptl.2019.03.015

Azizan, N. H., Mahmud, Z., & Rambli, A. (2020). Rasch rating scale item estimates using maximum likelihood approach: Effects of sample size on the accuracy and bias of the estimates. *International Journal of Advanced Science and Technology, 29*(4), 2526–2531.

Barbic, S. P., & Cano, S. J. (2016). The application of Rasch measurement theory to psychiatric clinical outcomes research: Commentary on\ldots Screening for depression in primary care. *BJPsych Bulletin, 40*(5), 243–244. https://doi.org/10.1192/pb.bp.115.052290

Bond, T. G., Fox, C. M., & Lacey, H. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences.* Routledge,.

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

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

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q 3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178–194.

Conaway, B. J., Saxon, T. F., & Woods, M. B. (2003). A comparison of the reading abilities of teacher candidates. *The Teacher Educator, 39*(1), 1–17. https://doi.org/10.1080/08878730309555326

Cowan, N. (2015). George Miller's magical number of immediate memory in retrospect: Observations on the faltering progression of science. *Psychological Review, 122*(3), 536–541. https://doi.org/10.1037/a0039035

Cunningham, A. E., & Stanovich, K. E. (2001). What reading does for the mind. *Journal of Direct Instruction, 1*(2), 137–149.

Dawes, J. (2008). Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5-Point, 7-Point and 10-Point Scales. *International Journal of Market Research, 50*(1), 61–104. https://doi.org/10.1177/147078530805000106

Deliany, Z., & Cahyono, B. Y. (2020). Metacognitive reading strategies awareness and metacognitive reading strategies use of EFL university students across gender. *Studies in English Language and Education, 7*(2), 421–437. https://doi.org/10.24815/siele.v7i2.17026

Dinçer, B., & Çilek, G. (2022). The Analysis of the Relation Between Metacognitive Awareness of Reading Strategies and Critical Thinking Attiude of Pre-Service Classroom Teachers. *International Journal of Progressive Education, 18*(2). https://doi.org/10.29329/ijpe.2022.431.4

Fisher, W. P. J. (2007). Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions, 21*(1), 1095.

Fitrisia, D., Tan, K.-E., & Yusuf, Y. Q. (2015). Investigating metacognitive awareness of reading strategies to strengthen students' performance in reading. *Comprehension, 30* (1), 15–30.

Hair, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research, 109*, 101–110. https://doi.org/10.1016/j.jbusres.2019.11.069

Halim, N. A., Satimin, O., Obaid, A., & Ghazali, A. S. (2022). *Relationship between Metacognitive Awareness of Reading Strategies and Reading Comprehension, 10,* 13.

Hall, L. A. (2005). Teachers and content area reading: Attitudes, beliefs and change. *Teaching and Teacher Education, 21*(4), 403–414. https://doi.org/10.1016/j.tate.2005.01.009

Iwai, Y. (2016). Promoting Strategic Readers: Insights of Preservice Teachers' Understanding of Metacognitive Reading Strategies. *International Journal for the Scholarship of Teaching and Learning, 10*(1), n1. https://doi.org/10.20429/ijsotl.2016.100104

Jaleel, S., & P, P. (2016). A study on the metacognitive awareness of secondary school students. *Universal Journal of Educational Research, 4*(1), 165–172. https://doi.org/10.13189/ujer.2016.040121

Jamieson, S. (2004). Likert scales: How to (ab) use them? *Medical Education, 38*(12), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Kallio, H., Virta, K., & Kallio, M. (2018). Modelling the components of metacognitive awareness. *International Journal of Educational Psychology, 7*(2), 94. https://doi.org/10.17583/ijep.2018.2789

In *Rasch Measurement*. In Khine, M. S. (Ed.), *Rasch Measurement*, (2020). https://doi.org/10.1007/978-981-15-1800-3

Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., Van Buuren, S., Van der Beek, A. J., & De Vet, H. C. (2014). Improving the individual work performance questionnaire using rasch analysis. *Journal of Applied Measurement, 15*(2), 160–175.

Lesley, M., Watson, P., & Elliot, S. (2007). School" reading and multiple texts: Examining the metacognitive development of secondary-level preservice teachers. *Journal of Adolescent & Adult Literacy, 51*(2), 150–162. https://doi.org/10.1598/JAAL.51.2.6

Leung, S.-O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research, 37*(4), 412–421. https://doi.org/10.1080/01488376.2011.580697

Levchyk, I., Chaikovska, H., Mazur, O., Adamska, Z., & Zakordonets, N. (2022). The Impact of Metacognitive Reading Strategies on Master Students' EFL Reading Proficiency and Academic Achievement. *Acta Paedagogica Vilnensia, 48*, 26–46. https://doi.org/10.15388/ActPaed.2022.48.2

Linacre, John M. (2021). *Winsteps® Rasch measurement computer program User's Guide.* Winsteps.com.,.

Linacre, J.M. (2022). *Winsteps® (Version 5.2.2) [Computer Software].* (5.2.2). Winsteps.com.

Mahdavi, J. N., & Tensfeldt, L. (2013). Untangling reading comprehension strategy instruction: Assisting struggling readers in the primary grades. *Preventing School Failure: Alternative Education for Children and Youth, 57*(2), 77–92. https://doi.org/10.1080/1045988X.2012.668576

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. https://doi.org/10.1037/h0043158

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249–259. https://doi.org/10.1037/0022-0663.94.2.249

Mokhtari, K., & Reichard, C. A. (2004). Investigating the strategic reading processes of first and second language readers in two different cultural contexts. *System, 32*(3), 379–394.

Nahar, J., & Mallik, S. (2022). METACOGNITIVE AWARENESS IN THE READING COMPREHENSION OF LITERARY TEXTS: AN EXPERIMENTAL STUDY BASED ON METACOGNITIVE READING STRATEGY. *European Journal of English Language Teaching, 7*(1). https://doi.org/10.46827/ejel.v7i1.4145

Oon, P. T., & Fan, X. (2017). Rasch analysis for psychometric improvement of science attitude rating scales. *International Journal of Science Education, 39*(6), 683–700. https://doi.org/10.1080/09500693.2017.1299951

Park, M., & Liu, X. (2019). An Investigation of Item Difficulties in Energy Aspects Across Biology, Chemistry, Environmental Science, and Physics. *Research in Science Education.* https://doi.org/10.1007/s11165-019-9819-y

Pesämaa, O., Zwikael, O., Hair, J. F., & Huemann, M. (2021). Publishing quantitative papers with rigor and transparency. *International Journal of Project Management, 39* (3), 217–222. https://doi.org/10.1016/j.ijproman.2021.03.001

Pesudovs, K., & Noble, B. A. (2005). Improving Subjective Scaling of Pain Using Rasch Analysis. *The Journal of Pain, 6*(9), 630–636. https://doi.org/10.1016/j.jpain.2005.04.001

Pimentel, J. L., & Pimentel, J. L. (2019). Some biases in Likert scaling usage and its correction. *International Journal of Science: Basic and Applied Research (IJSBAR), 45* (1), 183–191.

Pressley, M. (2008). Epilogue: What the future of reading research could be. In C. C. Block, & S. R. Parris (Eds.), *Comprehension instruction: Research-based best practices* (pp. 391–413). New York: Guilford Press.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: MESA Press,.

Rezat, S., Malik, S., & Leifeld, M. (2022). Scaffolding Close Reading of Mathematical Text in Pre-service Primary Teacher Education at the Tertiary Level: Design and Evaluation. *International Journal of Science and Mathematics Education.* https://doi.org/10.1007/s10763-022-10309-y

Saaty, T. L., & Ozdemir, M. S. (2003). Why the magic number seven plus or minus two. *Mathematical and Computer Modelling, 38*(3), 233–244. https://doi.org/10.1016/S0895-7177(03)90083-5

Sarimanah, E., Soeharto, S., Dewi, F. I., & Efendi, R. (2022). Investigating the relationship between students' reading performance, attitudes toward ICT, and economic ability. *Heliyon, 8*(6), Article e09794. https://doi.org/10.1016/j.heliyon.2022.e09794

Shah, S. M. A., & Asgher, D. T. (2022). *Reading Strategies of Male and Female Pakistani ESL Undergraduates in the Context of Reading Indigenized Academic Texts, 6*(3), 11.

Sheorey, R., & Mokhtari, K. (2001). Differences in the metacognitive awareness of reading strategies among native and non-native readers. *System, 29*(4), 431–449.

Skaar, H., Elvebakk, L., & Nilssen, J. H. (2018). Literature in decline? Differences in pre-service and in-service primary school teachers' reading experiences. *Teaching and Teacher Education, 69*, 312–323. https://doi.org/10.1016/j.tate.2017.10.019

Soeharto, S., & Csapó, B. (2022). Assessing Indonesian student inductive reasoning: Rasch analysis. *Thinking Skills and Creativity, 46*, Article 101132. https://doi.org/10.1016/j.tsc.2022.101132

Sulentic-Dowell, M. M., Beal, G. D., & Capraro, R. M. (2006). How do literacy experiences affect the teaching propensities of elementary pre-service teachers? *Reading Psychology, 27*(2–3), 235–255. https://doi.org/10.1080/02702710600642509

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Upadhyay, S. S. N., Brusnighan, S. M., & Folk, J. R. (2022). Readers accurately monitor their incidental word learning success during silent reading. *Acta Psychologica, 228*, Article 103653. https://doi.org/10.1016/j.actpsy.2022.103653

Van Gelderen, A., Schoonen, R., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic Knowledge, Processing Speed, and Metacognitive Knowledge in First-and Second-Language Reading Comprehension: A Componential Analysis. *Journal of Educational Psychology, 96*(1), 19–30.

Villanueva, J. M. (2022). Language profile, metacognitive reading strategies, and reading comprehension performance among college students. *Cogent Education, 9*(1), Article 2061683. https://doi.org/10.1080/2331186X.2022.2061683

Wigfield, A., Gladstone, J. R., & Turci, L. (2016). Beyond cognition: Reading motivation and reading comprehension. *Child Development Perspectives, 10*(3), 190–195. https://doi.org/10.1111/cdep.12184

Wilson, N. S., & Bai, H. (2010). The relationships and impact of teachers' metacognitive knowledge and pedagogical understandings of metacognition. *Metacognition and Learning, 5*(3), 269–288. https://doi.org/10.1007/s11409-010-9062-4

Wu, H., & Leung, S.-O. (2017). Can Likert scales be treated as interval scales?—A Simulation study. *Journal of Social Service Research, 43*(4), 527–532. https://doi.org/10.1080/01488376.2017.1329775

Yorulmaz, A., Uysal, H., & Çokçaliskan, H. (2021). Pre-service primary school teachers' metacognitive awareness and beliefs about mathematical problem solving. *JRAMathEdu (Journal of Research and Advances in Mathematics Education), 6*(3), 239–259. https://doi.org/10.23917/jramathedu.v6i3.14349

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.