Vol. (11), No. (2); 2025
ISSN : 2355-2069 (Print)
ISSN : 2656-5765 (Online)
Published by UIN Fatmawati Sukarno Bengkulu

DEVELOPING METACOGNITIVE AWARENESS READING QUESTIONNAIRE (MARQ) FOR READING COMPREHENSION; A PRELIMINARY STUDY

DWI INDRA APRILLIANDARI[1], RINDU HANDAYANI[2]
[1,2]*Universitas Muhammadiyah Bangka Belitung*
dwi.indraaprilliandari@unmuhbabel.ac.id, rindu.handayani@unmuhbabel.ac.id

DOI: http://dx.doi.org/10.29300/ling.v11i2.5370

Received: September 30th 2024 Accepted: January 6th 2025 Published: December 20th 2025

Abstract

This initial study seeks to create a Metacognitive understanding Reading Questionnaire (MARQ) to evaluate university students' understanding of metacognitive methods in reading comprehension. The MARQ emphasises critical tactics including preparation, monitoring, and evaluation, which are vital for enhancing reading comprehension, especially in academic and assessment environments such as the TOEFL. The study used a qualitative method to assess the initial usefulness of the MARQ, highlighting its potential to guide treatments and improve students' metacognitive awareness and reading ability. Three experts were asked to validate the questionnaire; one expert to validate a construct validation sheet and two experts to validate the content validity sheet. The statements in validation sheets used a Likert-scale from a range of Strongly Agree and Strongly Disagree. The results from the experts were then described in percentages to obtain the validity and reliability of the questionnaire. It is recommended that the tool undergo more improvement for wider use. Integrating metacognitive strategy training into the academic curriculum may augment learning results, enhance reading efficiency, and elevate overall performance in high-stakes reading assessments.

Keywords: MARQ, reading comprehension, TOEFL Reading

INTRODUCTION

Metacognition, a concept initially proposed by Flavell (Flavell, 1979), denotes an individual's knowledge and manipulation of their cognitive processes. In academic settings, metacognition is seen as essential for success as it enables learners to proficiently organise, monitor, and assess their cognitive processes (Baker & Brown, 1980). Metacognitive skills are particularly essential for university students, who must interact with intricate texts across various academic fields. Metacognitive awareness considerably enhances reading comprehension, a key ability. Students who possess an awareness of their cognitive processes can more effectively modify their techniques to accommodate various reading materials and academic tasks (khellab et al., 2022). Research repeatedly indicates that individuals possessing elevated metacognitive awareness generally achieve superior academic performance, since they can adjust their learning tactics to various circumstances and disciplines (Efklides, 2006). Nonetheless, a significant deficiency persists in the domain: whereas various tools are available to evaluate metacognition in general, few particularly target metacognitive awareness in reading among university students.

The academic achievement of university students is intimately linked to their effective application of metacognitive strategies, which play a crucial role in reading comprehension. University students often encounter complex, technical, or abstract materials, necessitating the use of sophisticated reading skills to synthesise information, draw inferences, and apply knowledge in many situations (Dhieb-Henia, 2003). Nevertheless, despite the acknowledged significance of metacognitive methods in reading, numerous students encounter difficulties in employing these strategies proficiently, especially in higher education settings where the requirements of academic reading are substantial (Banditvilai, 2020; Jake Follmer & Sperling, 2018) Consequently, there is an increasing interest in creating tools that can evaluate students' metacognitive awareness and facilitate treatments to improve reading comprehension. Studies indicate that individuals utilising metacognitive methods generally achieve superior results on reading comprehension assessments such as the TOEFL.

A study of language learners preparing for the TOEFL indicated that individuals who frequently employed metacognitive methods attained superior reading scores compared to their colleagues who depended exclusively on fundamental cognitive strategies like memorisation or decoding (Andersen, 2002). This discovery underscores the need to include metacognitive strategies into TOEFL preparation programs, as these strategies promote improved regulation of the reading process and augment understanding under time limitations. By enhancing their awareness of reading habits and self-regulating their strategies, students can elevate their performance on specific exam questions and their overall capacity to comprehend complicated academic books in English (Sun et al., 2024; Teng & Zhang, 2024). Consequently, metacognitive strategy training ought to be regarded as a fundamental element of successful TOEFL preparation. University students frequently face numerous obstacles in the reading comprehension segment of the TOEFL, many of which are intricately linked to deficiencies in their application of metacognitive methods.

A prevalent issue is the challenge of monitoring understanding, which is a fundamental aspect of metacognitive awareness (Manh Do & Le Thu Phan, 2021). Students may fail to identify instances of incomplete understanding of a book, resulting in an inadequate grasp of essential details or principal concepts. Insufficient self-awareness when reading may cause students to misinterpret questions or overlook essential information, hindering their ability to respond correctly. Nevertheless, current instruments often fail to address the distinctive metacognitive requirements of high-stakes evaluations, such as the TOEFL, especially in the Indonesian context. This preliminary study seeks to develop and validate the Metacognitive Awareness Reading Questionnaire (MARQ), an instrument intended to evaluate university students' metacognitive strategies during the pre-reading, while-reading, and post-reading phases of TOEFL-type reading tasks. This study aims to address the following enquiries: 1) To what extent does the MARQ demonstrate adequate content and construct validity as assessed by expert judgement?, 2) To to what extent does the MARQ demonstrate adequate item validity and internal consistency reliability when administered to university students?

This study aims to develop and initially validate the MARQ, an adaptable instrument for identifying students' metacognitive profiles in reading, hence guiding the design of strategy-based instruction and TOEFL preparation programs in Indonesian universities.

METHOD

Design

This study utilised a preliminary instrument development design to create and validate the Metacognitive Awareness Reading Questionnaire (MARQ) for TOEFL reading comprehension. The validation procedure integrated expert evaluation (content and construct validity) and empirical testing with student participants to assess item validity and internal consistency reliability.

Respondents

The study involved two groups of respondents. At first, three experts were engaged to assess the clarity, relevance, and theoretical coherence of the MARQ components. One expert specialised in educational psychology, while two others were professionals in reading comprehension and TOEFL preparation. They evaluated both the construct and content dimensions of the instrument. Secondly, thirty-seven university students from a private institution in Bangka Belitung served as pilot respondents. The participants finalised the 39-item MARQ, yielding empirical data for item analysis and reliability assessment.

Instruments

Experts validation sheets

Two validation sheets were created for the experts. A construct validity document containing fifteen criteria pertaining to the correspondence between each MARQ item and the theoretical components of metacognitive awareness (planning, monitoring, and evaluation). A content validity document consisting of eleven criteria that emphasise the significance, utility, clarity, and pertinence of the items for TOEFL reading settings and student requirements. Both questionnaires employed a 4-point Likert scale, ranging from "Strongly Agree" (4) to "Strongly Disagree" (1). The experts' responses were subsequently transformed into percentages and classified into validity categories, namely very valid, valid, reasonably valid, invalid.

Metacognitive Awareness Reading Questionnaire (MARQ)

The MARQ comprises 39 items that assess students' metacognitive strategies across three phases of reading. Pre-reading (Planning): Items P1–P7 concentrate on establishing goals, previewing the text, activating prior knowledge, and formulating strategies prior to reading. Items P8–P22 focus on monitoring comprehension, assessing understanding, modifying strategies, and addressing challenges encountered during reading. Post-reading (Evaluation): Items P23–P39 focus on assessing comprehension, reflecting on the effectiveness of the strategy, and reviewing performance subsequent to reading. Items were evaluated using a 5-point Likert scale, with responses ranging from 1 (“Never true”) to 5 (“Always true”).

Procedures

Firstly, the researchers primarily modified the metacognitive strategy categories established by Zhang and Seepho (2013), which revised the previous framework of O'Malley and Chamot (1990) for academic reading contexts. The questionnaire was structured around three primary categories: Pre-reading (planning), While-reading (monitoring), and Post-reading (evaluation). The original items were adapted for TOEFL reading tasks, incorporating local content pertinent to Indonesian university students in Bangka Belitung, especially in the pre-reading items to contextualise goal setting and activate prior knowledge. Secondly, the draft MARQ items were converted into construct and content validation sheets and submitted to three experts. The construct validation sheet included fifteen statements that assessed the extent to which the MARQ items represented essential metacognitive dimensions. The content validation sheet included eleven statements that addressed relevance to both academic and non-academic texts, clarity of language, alignment with student needs, and appropriateness for different levels of English proficiency. Experts evaluated each criterion using a 4-point scale and were able to offer qualitative feedback for revisions. Thirdly, in accordance with expert recommendations and percentage scores, the wording and focus of several items were refined. The expert validation results demonstrated significant agreement across most criteria, classified as "extremely valid" and "very good" trustworthiness, indicating that the MARQ items were both theoretically robust and contextually suitable. The revised MARQ was administered to 37 student respondents. All 39 items were completed in a single session. The total scores for each item and the overall scale were utilised to assess item validity via item–total correlations and scale reliability through Cronbach's alpha.

Data Analysis

Data from the expert validation sheets were analysed using descriptive statistics. The scores for each criterion were summed, converted into percentages, and interpreted using the following classification: $p = \sum x \sum xi \times 100\%$. Two steps were taken into account: 1) item–total correlation, 39 items were analysed for correlation with the total MARQ score through Pearson's product–moment correlation method. The coefficients (R_{count}) were compared to the critical value (R_{table}) at a significance level of 0.05 for a sample size of 37. Items exhibiting correlation coefficients exceeding 0.325 were deemed valid, whereas those with coefficients at or below 0.325 were regarded as lacking adequate discriminative power, necessitating revision or removal. 2)Cronbach's alpha, the internal consistency of the MARQ, was evaluated using Cronbach's alpha coefficient. If the computation for all 39 items produced an alpha of approximately 0.91, it would be interpreted as excellent internal consistency for educational and psychological instruments.

FINDINGS

The findings from expert judgements and student feedback yield converging evidence that the MARQ is a psychometrically promising tool, while also highlighting numerous issues that necessitate adjustments. The results are categorised below into expert validation and item-level validity.

Experts validation results

Content Validation

The content validation performed by the two expert reviewers establishes a solid framework for asserting that the MARQ is theoretically and contextually suitable for evaluating metacognitive awareness in TOEFL-like reading tasks. Both experts scored the instrument at 91%, categorising it as "extremely valid" and indicating "very good" reliability for classroom and research applications. Experts concurred that the questionnaire comprehensively addresses both academic and non-academic reading texts, ensuring that the assertions are applicable to the wide range of materials typically encountered by university students. Elevated grades (75–100%) for criteria concerning the clarity of the test's goal indicate that the items effectively convey the rationale for administering the reading test and the anticipated level of student engagement, both of which are essential for fostering strategic and self-regulated reading. Tables 1 and 2 below describe the results from content validity experts 1 and 2.

Table 1. Result from Content Validity Expert 1

Aspect	Criteria	%	Validity	Trustworthiness
Content Validity	Statement items with readings for academics are very relevant	100	Extremely Valid	Very Good
	Statement items with non-academic English text are very relevant	75	Extremely Valid	Very Good
	The statement item contains information about the purpose of taking the test.	100	Extremely Valid	Very Good
	There is a statement to direct students to be able to develop strategies for answering test questions.	100	Extremely Valid	Very Good
	There are statements to direct students to be able to guess the main idea of the reading.	100	Extremely Valid	Very Good
	There is a statement to direct students to be able to organize their time in answering test questions.	100	Extremely Valid	Very Good
	Statements are made following the student's condition of knowledge or prior knowledge.	100	Extremely Valid	Very Good
	Statement items can guide students to answer test questions with focus and full consideration.	100	Extremely Valid	Very Good
	Statement items can guide students to be able to review and reflect on attitudes toward their abilities after working on test questions.	100	Extremely Valid	Very Good
	The choice of words or diction in statement items is easy for students to understand.	75	Extremely Valid	Very Good
	Questionnaires can be given to all levels of English proficiency.	50	Fairly Valid	Fair
TOTAL SCORE		91	Extremely Valid	Very Good

Table 2. Result from Content Validity Expert 2

Aspect	Criteria	%	Validity	Trustworthiness
	Statement items with readings for academics are very relevant	100	Extremely Valid	Very Good
	Statement items with non-academic English text are very relevant	75	Extremely Valid	Very Good
	The statement item contains information about the purpose of taking the test.	100	Extremely Valid	Very Good
	There is a statement to direct students to be able to develop strategies for answering test questions.	100	Extremely Valid	Very Good

Content Validity	There are statements to direct students to be able to guess the main idea of the reading.	100	Extremely Valid	Very Good
	There is a statement to direct students to be able to organize their time in answering test questions.	100	Extremely Valid	Very Good
	Statements are made following the student's condition of knowledge or prior knowledge.	100	Extremely Valid	Very Good
	Statement items can guide students to answer test questions with focus and full consideration.	100	Extremely Valid	Very Good
	Statement items can guide students to be able to review and reflect on attitudes toward their abilities after working on test questions.	100	Extremely Valid	Very Good
	The choice of words or diction in statement items is easy for students to understand.	75	Extremely Valid	Very Good
	Questionnaires can be given to all levels of English proficiency.	50	Fairly Valid	Fair
TOTAL SCORE		91	Extremely Valid	Very Good

The experts assessed the items as highly helpful in guiding students toward fundamental metacognitive strategies, including identifying primary ideas, managing time, and evaluating performance after completing a set of questions. The elevated percentages suggest that the MARQ not only delineates general learning habits but also addresses behaviours pertinent to high-stakes reading contexts such as the TOEFL. Experts emphasise that a notable strength is the congruence between item phrasing and students' prior knowledge, indicating that most statements are well calibrated to topic familiarity and cognitive challenge. The primary limitation in the content validation is the criterion "questionnaires can be administered to all levels of English proficiency," which both experts assessed at merely 50% ("fairly valid"). This diminished score suggests that certain items may include vocabulary or grammatical patterns that are difficult for students with very poor proficiency, yet are appropriate for standard college EFL learners. Therefore, although the MARQ is well endorsed for intermediate and advanced university students, further refinement or scaffolding is required before its implementation with beginner or remedial groups.

Construct Validation

A third expert's construct validation assesses the extent to which the MARQ aligns with the theoretical framework of metacognitive reading, including the phases of planning, monitoring, and evaluation. The overall score of 85% categorises the instrument as "extremely valid" with "very good" reliability, suggesting that the expert assessed the questionnaire as mainly aligned with current models of metacognition in reading. In the planning (pre-reading) dimension, multiple criteria achieved the highest rating (100%), encompassing elements that assist students in generating preliminary ideas about the text, clarifying their reading objectives, organising sections of the reading, anticipating significant components, formulating strategies for responding to questions, and augmenting the text with their own knowledge. The results indicate that the planning elements together promote students in establishing goals, drawing on prior knowledge, and anticipating the text's structure, which are widely acknowledged as characteristics of effective metacognitive preparation. The table is presented below.

Table 3. Result from Construct Validity Expert

Aspect	Criteria	%	Validity	Trustworthiness
	Planning (Pre-reading)			
Construct Validity	The question items developed aim to find out the initial idea of the text.	100	Extremely Valid	Very Good
	The question items developed are able to provide an understanding of the purpose of the reading.	75	Extremely Valid	Very Good
	The question items developed can be planned into reading subsections.	100	Extremely Valid	Very Good
	Question items can help students predict the text into several important parts of the reading.	100	Extremely Valid	Very Good
	Question items can help students create strategies in solving questions in reading.	100	Extremely Valid	Very Good
	Question items can help students elaborate the text with their knowledge.	100	Extremely Valid	Very Good
	Question items can help students focus on completing reading assignments.	50	Fairly Valid	Fair
	Question items can help students modify reading strategies.	75	Extremely Valid	Very Good
	Monitoring (While Reading)			
	Question items can help students monitor understanding, accuracy and appropriateness in the reading process and text.	75	Extremely Valid	Very Good
	Question items can help students observe their abilities and difficulties in reading.	100	Extremely Valid	Very Good
	Question items can help students connect the reading strategies they have learned and can be applied to reading tests.	75	Extremely Valid	Very Good
	Question items can help students choose alternative strategies if they feel they are unable to answer.	75	Extremely Valid	Very Good
	Evaluating (Post-reading)			
	Question items can help students assess their ability to understand reading.	100	Extremely Valid	Very Good
	Question items can help students evaluate their ability to understand reading and the strategies used to understand reading.	75	Extremely Valid	Very Good
	Question items can help students reflect on their ability to understand reading and the learning process they have gone through.	75	Extremely Valid	Very Good
TOTAL SCORE		85	Extremely Valid	Very Good

One criterion in the planning dimension, however, received a score of only 50% and was classified as “fairly valid”: the extent to which items assist students in sustaining focus during reading assignments. This suggests that, according to the expert, the existing planning elements inadequately encompass the attentional and motivational aspects necessary for maintaining engagement with a text, despite effectively addressing goal setting and strategy design. The consequence for future revisions is that one or two elements may directly pertain to attention management, such as combating distractions or refocusing after lapses, to enhance this component of the construct. In the monitoring (while-reading) dimension, all criteria received

ratings between 75% and 100%, indicating that the expert deemed the items useful in assisting students with comprehension monitoring, identifying challenges, and connecting learnt tactics to current assessment requirements. In the evaluation (post-reading) dimension, criteria related to assessing understanding, evaluating the efficacy of techniques, and reflecting on one's learning experience likewise achieved scores between 75–100%, categorising them within the "extremely valid" range. The consistently elevated results for monitoring and assessment suggest that, theoretically, the MARQ effectively encapsulates the dynamic processes of assessing comprehension and reflecting on performance, rather than merely static knowledge of techniques. The construct validation supports the three-phase framework of the MARQ, emphasising the need to enhance focus-maintenance elements during the planning phase.

Empirical Item Validity

Total Validity

Empirical validation with 37 university students revealed item-level evidence about the operational functionality of the MARQ. Figure 1 presents a succinct visual representation of the overall validity and reliability of the MARQ, based on responses from 37 university students. The pie chart is segmented into two portions, representing the ratios of valid to invalid items for the total of 39 statements in the questionnaire. The blue segment indicates that 33 items (84.6% of the instrument) met the minimal item–total correlation threshold of 0.325 ($\alpha = 0.05$) and are thus categorised as reliable indicators of metacognitive awareness in reading. The red section, in comparison, denotes the remaining six items (15.4%) that did not meet this criterion and are deemed invalid or problematic. The clear visual differentiation between the blue and red regions enables readers to rapidly comprehend that the majority of MARQ items operate effectively, whilst a minor subset necessitates improvement.

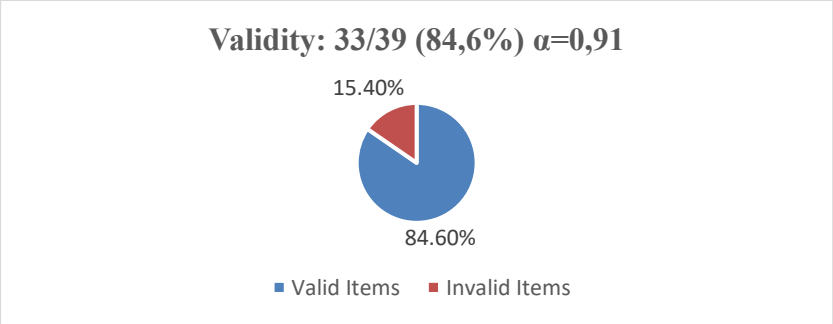

Figure 1. Total Validity

The "Validity: 33/39 (84.6%) $\alpha = 0.91$," reinforces this interpretation by combining information on both item validity and scale reliability in a single statement. The first part of the title reiterates that 33 out of 39 items are valid, matching the proportions depicted in the chart, while the second part reports the Cronbach's alpha coefficient of 0.91. This reliability value indicates excellent internal consistency for educational and psychological instruments, meaning that students responded to the items in a highly consistent manner and that the MARQ operates coherently as a single scale. In other words, even though a small number of items show weak correlations with the total score, their presence does not undermine the overall

reliability of the questionnaire. In the larger findings section, Figure 1 presents a comprehensive summary that precedes and enhances the analyses of item–total correlations presented in the following tables and figures. This indicates to readers that the instrument demonstrates psychometric potential—most items are valid, and the overall scale exhibits good reliability—while also necessitating refinement of the six failing items identified in the empirical item-level analysis.

Metacognitive Phase Validity

To transcend a singular overall validity coefficient and identify the instrument's strengths and weaknesses, the findings also analyse item performance inside each metacognitive phase of the MARQ. Figure 2 provides a phase-level overview of the validity of MARQ items across the three metacognitive stages: planning (pre-reading), monitoring (while reading), and evaluation (post-reading). The blue bars illustrate the quantity of valid items in each phase, whilst the red bars signify invalid items that failed to surpass the item–total correlation threshold of 0.325. Above each cluster of bars, the mean item–total correlation (Mean r) is presented, offering supplementary insight into the overall efficacy of item performance within that phase. The picture below is the description of the metacognitive phase.

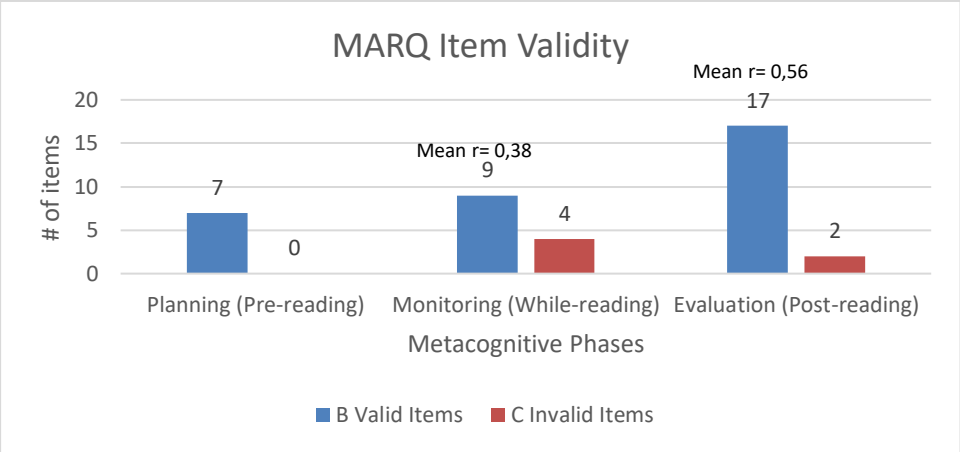

Figure 2. Metacognitive Phases Validity

During the planning phase, all seven items are legitimate, with none below the cut-off, yielding a mean correlation of 0.56, the highest among the three stages. This pattern demonstrates that planning statements—such as previewing texts, establishing reading objectives, and activating past knowledge—are consistently perceived by students and are significantly correlated with their overall metacognitive awareness. The lack of invalid items indicates that the phrasing and conceptual emphasis of the planning items are well aligned with both the theoretical framework and the students' reading experiences. By contrast, the monitoring phase shows a more mixed profile, with nine valid items and four invalid items, and a notably lower mean correlation of 0.38. This indicates that although most monitoring items function acceptably, several statements do not discriminate well between higher- and lower-metacognitive readers, likely reflecting difficulties with self-reporting real-time comprehension checks or ambiguity in their phrasing. The evaluation phase falls between these two extremes: 17 of 19 items are valid, two are invalid, and the mean correlation is 0.53, indicating generally strong functioning, with only a small number of underperforming items.

Overall, Figure 2 shows that the MARQ is exceptionally robust in assessing planning and evaluation. In contrast, four monitoring items and two evaluation items require targeted revision to strengthen phase-level balance and construct coverage.

Item Correlation

To enhance phase-level analysis and accurately determine whether statements reinforce or undermine the instrument, the subsequent findings focus on the validity of each item individually. Figure 3 presents a comprehensive item-by-item representation of the MARQ item-total correlations (Rcount) for all 39 assertions. Each blue or red bar represents an individual item (P1–P39) on the horizontal axis, and the vertical axis indicates the correlation coefficient, which ranges from approximately 0.00 to 0.90. The horizontal orange line indicates the critical value of Rtabel = 0.325 (α = 0.05, n = 37), which serves as the threshold for classifying outcomes as legitimate or invalid. Bars that exceed this line signify items with adequate discrimination, while bars that fall below suggest items with an insufficient correlation to the total MARQ score. The figure can be seen below.

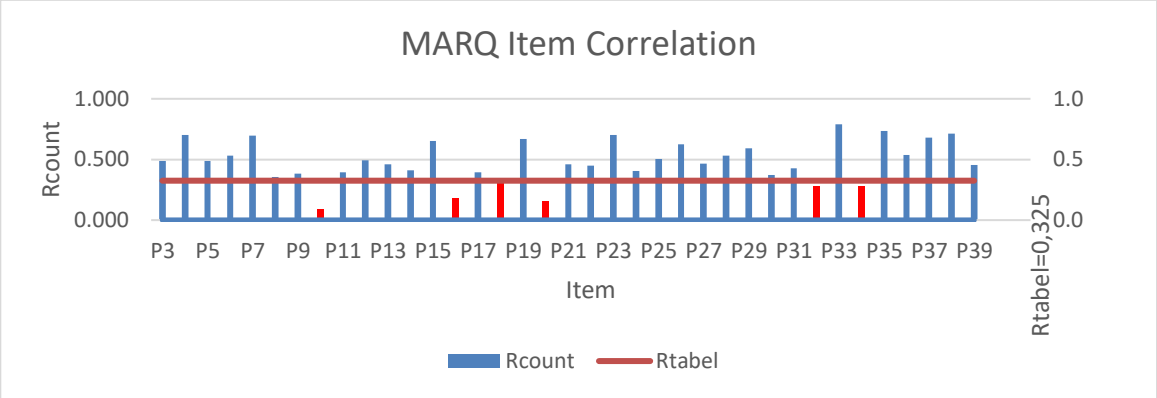

Figure 3. Item Correlation

The majority of items are represented as blue bars significantly beyond the Rtabel line, visually corroborating that 33 out of 39 items satisfy the validity requirement and reinforcing the previous numerical conclusion that 84.6% of the instrument is psychometrically robust. Six red bars—representing items P10, P16, P18, P20, P32, and P34—fall beneath the threshold, rendering them readily recognised as problematic items necessitating adjustment or possible elimination. P10 exhibits the lowest value, indicating a poor correlation of 0.0871, whereas items like P33 demonstrate significantly higher values nearing 0.80, signifying exceptional discriminative ability within the same scale.

DISCUSSION

This initial study aimed to develop and evaluate the Metacognitive Awareness Reading Questionnaire (MARQ) for application in TOEFL-type reading comprehension assessments among university students. The amalgamated expert and empirical evidence suggests that the MARQ exhibits psychometric potential, while also highlighting particular domains necessitating enhancement. Validation of content by two experts yielded ratings of 91% for each reviewer, categorising the instrument as "extremely valid" and affirming that the items are highly pertinent to both academic and non-academic reading texts, effectively convey test

objectives, and guide students in essential strategies such as identifying main ideas, time management, and performance reflection. A third expert's construct validation resulted in a total score of 85%, categorising it within the “extremely valid” range. This assessment indicated that the items effectively represent the three phases of metacognitive reading—planning, monitoring, and evaluation—though one planning criterion concerning sustained focus was deemed only “fairly valid.” The empirical data support and nuance these expert opinions. Item–total correlation analysis with 37 students found that 33 of 39 items (84.6%) met the necessary value of 0.325 and hence function as legitimate indicators of metacognitive awareness, with a mean correlation of around 0.48. All seven planning measures and most evaluation items exhibited substantial relationships, showing that students consistently reported pre-reading and post-reading behaviours in ways that match with overall metacognitive ability. This pattern resonates with previous research demonstrating that effective readers intentionally set goals, preview texts, and evaluate the success of their strategies after reading, which enhances comprehension of dense academic materials (Nurdianingsih, 2021; Pandiangan et al., 2021; Romadhon, 2024).

In contrast, four monitoring items (P10, P16, P18, P20) and two evaluation items (P32, P34) did not meet the validity threshold, and the monitoring phase exhibited the lowest mean correlation, indicating that online self-monitoring is both conceptually intricate and more challenging to quantify via self-report. This corresponds with previous research indicating that students frequently have difficulties in recognising comprehension failures and modifying tactics during reading, despite their ability to articulate planning and evaluation activities more readily (Royanto, 2012; Tajalli & Satari, 2013). The elevated Cronbach’s alpha of 0.91 signifies exceptional internal consistency, affirming that the MARQ functions cohesively as a scale, although the existence of some weak items. The discussion leads to a balanced conclusion: the MARQ currently offers a valid and reliable assessment of students’ metacognitive awareness, particularly in planning and evaluation. However, a focused revision of six items, especially those related to monitoring, will enhance construct coverage and measurement accuracy in subsequent validation phases.

The findings indicate that, in this cohort, metacognitive awareness in reading is more reliably developed during pre-reading planning and post-reading evaluation than during real-time monitoring, where pupils demonstrate significant deficiencies. The robust expert validity indices (91% and 85%) and good internal consistency ($\alpha = 0.91$) suggest that the MARQ has substantial potential as a diagnostic instrument for distinguishing between well-develop d and underdeveloped aspects of metacognitive regulation. The small number of non-functioning items, particularly those intended for self-monitoring, indicates conceptual and practical issues that must be resolved in the next version of the instrument, rather than detracting from its overall measurement quality. The MARQ not only provides a robust initial assessment of students' metacognitive awareness but also outlines a clear framework for pedagogical intervention and subsequent scale enhancement, reinforcing the notion that focused strategy instruction should emphasise strengthening monitoring processes to augment students' current planning and evaluation capabilities (Zalha et al., 2020).

Advancing MARQ-type research is essential given the intricate nature of metacognition and reading (Rosnaeni et al., 2020; Saif et al., 2021; Soto et al., 2019). Initial instruments across

various domains, such as HOTS assessments and science-literacy tests, exhibit a consistent mix of valid and invalid items, despite acceptable reliability. This underscores the necessity of iterative refinement as an obligatory process rather than a discretionary one . The results from MARQ and reading literacy assessments indicate that monitoring and other higher-order processes are consistently the weakest and least accurately evaluated aspects, despite their significance to advanced academic performance. (Jing et al., 2025; Masitoh et al., 2023; Maxnun et al., 2024). Future research should implement multi-method validation that integrates Classical Test Theory with Rasch or other Item Response Theory models, factor analysis, and, when feasible, predictive correlations with actual performance, as reliance on a single analytical perspective may conceal misfitting items and overstate confidence in defective scales (Maxnun et al., 2024; McNamara, 2017; Syam & Ermawati, 2024).

CONCLUSION

This study created and initially validated the Metacognitive Awareness Reading Questionnaire (MARQ) for application in TOEFL-type reading tasks among university students. Expert assessment demonstrated robust content and construct validity (91% and 85%), indicating that the items are theoretically sound, contextually relevant, and consistent with the three phases of metacognitive reading: planning, monitoring, and evaluation. Empirical testing with 37 students demonstrated that 33 of 39 items (84.6%) surpassed the item-total correlation threshold, and the overall scale exhibited exceptional internal consistency ($\alpha = 0.91$), validating that the MARQ operates as a cohesive and dependable tool. Simultaneously, six items—primarily in the monitoring phase—exhibited weak correlations and necessitate focused adjustment, suggesting that online comprehension monitoring is the most difficult element to assess. The MARQ serves as a valuable instrument for assessing students' metacognitive strengths and weaknesses in reading and guiding strategy-based instruction. Future research should address the identified issues with certain items, validate the factor structure using larger samples, and explore the correlation between MARQ scores and actual reading performance in high-stakes assessments.

REFERENCES

Andersen, N. J. (2002). *The Role of Metacognition in Second Language Teaching and Learning. ERIC Digest. ERIC Clearinghouse on Languages and Linguistics*. http://www.cal.org/ericcll/DIGEST.

Baker, L., & Brown, A. L. (1980). *CENTER FOR THE STUDY OF READING .METACOGNITIVE SKILLS AND READING*. Longman, in press.

Banditvilai, C. (2020). The Effectiveness of Reading Strategies on Reading Comprehension. *International Journal of Social Science and Humanity*, *10*(2), 46–50.

Dhieb-Henia, N. (2003). Evaluating the effectiveness of metacognitive strategy training for reading research articles in an ESP context. *English for Specific Purposes*, *22*(4), 387–417. https://doi.org/10.1016/S0889-4906(03)00017-6

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, *1*(1), 3–14. https://doi.org/10.1016/j.edurev.2005.11.001

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

Jake Follmer, D., & Sperling, R. A. (2018). Interactions between reader and text: Contributions of cognitive processes, strategy use, and text cohesion to comprehension of expository science text. *Learning and Individual Differences*, *67*, 177–187. https://doi.org/10.1016/j.lindif.2018.08.005

Jing, W., Rahman, A. B. A., & Xiao, X. (2025). The Mediating Role of Students' Metacognitive Awareness between Teachers' Reading Strategies and Students' Reading Performance. *Journal of Curriculum and Teaching*, *14*(3), 292–307. https://doi.org/10.5430/jct.v14n3p292

khellab, F., Demirel, Ö., & Mohammadzadeh, B. (2022). Effect of Teaching Metacognitive Reading Strategies on Reading Comprehension of Engineering Students. *SAGE Open*, *12*(4). https://doi.org/10.1177/21582440221138069

Manh Do, H., & Le Thu Phan, H. (2021). Metacognitive Awareness of Reading Strategies on Second Language Vietnamese Undergraduates. *Arab World English Journal*, *12*(1), 90–112. https://doi.org/10.24093/awej/vol12no1.7

Masitoh, F., Fitriyah, I., & Mukminatien, N. (2023). DEVELOPING READING ASSESSMENT INSTRUMENT FOR INTERMEDIATE EFL LEARNERS: AZWAR MODEL. *English Review: Journal of English Education*, *11*(3). https://doi.org/10.25134/erjee.v11i3.7461

Maxnun, L., Kristiani, K., & Sulistyaningrum, C. D. (2024). Development of hots-based cognitive assessment instruments: ADDIE model. *Journal of Education and Learning*, *18*(2), 489–498. https://doi.org/10.11591/edulearn.v18i2.21079

McNamara, D. S. (2017). Self-Explanation and Reading Strategy Training (SERT) Improves Low-Knowledge Students' Science Course Performance. *Discourse Processes*, *54*(7), 479–492. https://doi.org/10.1080/0163853X.2015.1101328

Nurdianingsih, F. (2021). *TEACHERS' STRATEGIES IN TEACHING READING COMPREHENSION*. *4*(2).

O'Malley, J. M., & Chamot, A. U. (1990). *Learning Strategies in Second Language Acquisition* (Vol. 1).

Pandiangan, T., Parulian, S., Gaol, L., & Saragih, E. (2021). READING TEACHING STRATEGIES APPLIED BY ENGLISH TEACHERS IN SENIOR HIGH SCHOOL. *Jurnal Pendidikan LLDIKTI Wilayah 1 (JUDIK)*, *1*(2), 72–81.

Romadhon, R. (2024). Enhancing Reading Comprehension Among Informatics Management Students through the Grammarian Android Game: A Classroom Action Research. *EBONY: Journal of English Language Teaching, Linguistics, and Literature*, *4*(2), 18–28. https://doi.org/10.37304/ebony.v4i2.13944

Rosnaeni, R., Dj, M. Z., & Nur, H. (2020). Students' Metacognitive Awareness and Reading Comprehension of Narrative Texts. *IJEE (INDONESIAN JOURNAL OF ENGLISH EDUCATION)*, 74–86. https://doi.org/10.15408/ijee.v7i1.17027

Royanto, L. R. (2012). The Effect of An Intervention Program based on Scaffolding to Improve Metacognitive Strategies in Reading: A Study of Year 3 Elementary School Students in Jakarta. *Procedia - Social and Behavioral Sciences*, *69*, 1601–1609. https://doi.org/10.1016/j.sbspro.2012.12.105

Saif, S., Ma, J., May, L., & Cheng, L. (2021). Complexity of test preparation across three contexts: case studies from Australia, Iran and China. *Assessment in Education: Principles, Policy and Practice*, *28*(1), 37–57. https://doi.org/10.1080/0969594X.2019.1700211

Soto, C., Gutiérrez de Blume, A. P., Jacovina, M., McNamara, D., Benson, N., Riffo, B., & Kruk, R. (2019). Reading comprehension and metacognition: The importance of inferential skills. *Cogent Education*, *6*(1). https://doi.org/10.1080/2331186X.2019.1565067

Sun, Q., Zhang, L. J., & Carter, S. (2024). English as a foreign language learners' metacognitive experiences and writing development: Unraveling the process of learning EFL writing. *Learning and Individual Differences*, *115*. https://doi.org/10.1016/j.lindif.2024.102540

Syam, A. T., & Ermawati, E. (2024). The ADDIE Model on the Development English Materials for the Constitutional Law Study Program. *Language Circle: Journal of Language and Literature*, *18*(2). http://journal.unnes.ac.id/nju/index.php/lc

Tajalli, P., & Satari, S. (2013). Effectiveness of Metacognitive Strategies on Reading Skills of Students with Hearing Disorders. *Procedia - Social and Behavioral Sciences*, *84*, 139–143. https://doi.org/10.1016/j.sbspro.2013.06.524

Teng, M. F., & Zhang, L. J. (2024). Validating a new method for assessing young foreign language learners' metacognitive knowledge. *Research Methods in Applied Linguistics*, *3*(2). https://doi.org/10.1016/j.rmal.2024.100121

Zalha, F. B., Alfiatunnur, A., & Kamil, C. A. (2020). Strategies in Dealing With the Reading Section of "Toefl Prediction": A Case of Aceh EFL Learners. *IJEE (Indonesian Journal of English Education)*, *7*(2), 2020. https://doi.org/10.15408/ijee.v7i1.17622

Zhang, L., & Seepho, S. (2013). Metacognitive Strategy Use and Academic Reading Achievement: Insights from a Chinese Context. In *Electronic Journal of Foreign Language Teaching* (Vol. 10, Issue 1). http://e-flt.nus.edu.sg/