P-ISSN: 2541-6960; E-ISSN: 2549-8754
Yupa: Historical Studies Journal
Vol. 8 No. 1, 2024 (24-34)
http://jurnal.fkip.unmul.ac.id/index.php/yupa

Development of E-Assessment Instruments for Assessing Metacognition Skills of Students in The Research Methodology Course

Nurnaningsih[1], Amrin[2]
[12]STKIP Taman Siswa Bima, Indonesia
[1]nurnaningsih1988@gmail.com, [2]amrinbinyunus1993@gmail.com

Submitted	Revised	Accepted
29/12/2023	23/01/2024	25/01/2024

Abstract This research is focused on the creation of a valid and dependable tool designed for the assessment of students' metacognitive abilities within the context of the Research Methodology course. The research methodology employed for this research follows the Research and Development (R&D) framework outlined by Mardapi, which encompasses ten distinct phases. The participants involved in this research comprised four lecturers responsible for instructing research methodology courses and a total of 61 students who were enrolled in these courses. The data collected for this research consisted of quantitative data acquired through expert validation questionnaires and trial instruments. Data analysis was conducted employing quantitative techniques, specifically utilizing Microsoft Excel and employing the Item Response Theory (IRT) Politomus data analysis approach within the R programming environment. The research outcomes indicated that the instrument employed to evaluate students' metacognitive skills in the research methodology course achieved a valid status as per expert evaluations, meeting the criteria for goodness. It was found to be valid in terms of the response distribution across all 24 items, and collectively, the instrument items were deemed capable of offering insights into the state of the test participants (respondents) by more than 85%. Moreover, the instrument demonstrated a very high level of reliability as 0,94.

Keywords: Metacognition, Instruments, IRT, polytomus

Abstrak *Penelitian ini bertujuan untuk mengembangkan instrumen yang dapat mengukur keterampilan metakognisi mahasiswa dalam mengikuti matakuliah Metodologi Penelitian yang valid dan reliabel. Jenis penelitian adalah R & D dari Mardapi yang terdiri dari 10 langkah. Subjek penelitian dalam penelitian ini adalah 4 dosen yang mengajar matakuliah metodelogi penelitian dan mahasiswa yang mengikuti matakuliah tersebut sebanyak 61 mahasiswa. Data yang dikumpulkan berjenis data kuantitatif dengan menggunakan instrumen angket validasi ahli, dan instrumen uji coba. Analisis data menggunakan analisis Kuantitatif menggunakan bantuan Microsoft excel dan IRT data Politomus menggunakan program R. Hasil penelitian menunjukkan bahwa instrument penilaian keterampilan metakognisi mahasiswa pada matakuliah metodelogi penelitian memenuhi kategori valid menurut para ahli dengan kriteria baik, valid secara proporsi pemilihan jawaban pada 24 butir, serta keseluruhan butir instrumen mampu memberikan informasi tentang keadaan peserta tes (responden) lebih dari 85%. Sedangkan reliabilitas instrumen sangat tinggi dengan skor Alpha Cronbach sebesar 0,94. .*

Kata kunci *:* Metakognisi, Instrumen, IRT, Politomus

INTRODUCTION

Metacognition skills are of paramount importance for students, particularly in the context of the 21st century. These skills are deemed essential for the cultivation of *soft skills*, as the development of *soft skills* is inseparable from the acquisition of metacognitive abilities. Metacognition, in itself, plays a pivotal role in enhancing an individual's insight and problem-solving acumen (Mitsea et al., 2021). Furthermore, metacognition skills exhibit a robust correlation with students' critical thinking abilities (Diella & Ardiansyah, 2017). Hence, it is imperative to prioritize the development and enhancement of students' metacognitive skills (Huda et al., 2021).

Metacognition is employed to denote an individual's consciousness of their own cognitive processes, particularly those closely associated with the act of thinking (Kundre & Pratini, 2019). It encapsulates the capacity to comprehend and manage one's own thought processes, thereby empowering individuals to arrive at improved and more efficient decisions when undertaking tasks. Metacognition skills can help students in constructing their understanding of abstract things, but need to be trained again through teacher assistance in applying the right strategies (Ramadhan & Pratana, 2020). Teacher assistance with the right strategy can be implemented if the measurement of students' metacognition can be clearly measured. Therefore, the existence of a valid and reliable instrument can reveal the ability of metacognition in students.

As of September 4, 2023, outcomes from interviews conducted with instructors responsible for teaching Research Methods revealed that certain students, particularly those with low to moderate academic abilities, encountered challenges in identifying the research problem they intended to investigate. While a good title must depart from a problem. This shows that the process of cognition, critical thinking and the ability to construct abstract to concrete understanding is still limited, so a strategy in research is needed.

The Research Methodology course holds significant importance for undergraduate students, serving as a foundational resource for research conduct. Nevertheless, several students frequently encounter challenges when it comes to comprehending and applying the concepts covered in this course. One of the factors that plays a pivotal role in determining students' success in this course is the level of their metacognition skills. Schraw, Jacob and Miller stated that bringing metacognition invites students to be able to regulate themselves towards planning, monitoring, and evaluation (Hong et al., 2015) and it is very much needed in research methodology courses. Therefore, the development of instruments that can measure students' metacognition skills in participating in Research Methodology courses is very important. This tool can assist educators in assessing students' capabilities and offering constructive feedback to aid in the enhancement of their metacognition skills. Furthermore, this instrument can be valuable for students themselves, enabling them to identify both their strengths and weaknesses in terms of metacognition skills. Consequently, this self-awareness can empower students to make necessary improvements and ultimately enhance their overall learning outcomes.

In light of the aforementioned explanation, the author's intention is to carry out a research research titled "Development Of E-Assessment Instruments For Assessing Metacognition Skills Of Students In The Research Methodology Course." This metacognition skills assessment instrument is packaged in online form (g-form) for easy access, saving paper usage, and faster to analyse the results. From this title, research

problems can be formulated, such as how is the validity and reliability of the developed metacognition assessment instrument?

METHODS

This research follows the Research and Development (R&D) approach introduced by Mardapi, which involves ten distinct developmental phases. The 10 development steps are 1) determining instrument specifications, 2) writing instruments, 3) determining instrument scales, 4) determining scoring systems, 5) reviewing instruments, 6) conducting trials, 7) analysing instruments, 8) assembling instruments, 9) carrying out measurements, and 10) interpreting measurement results (Mardapi, 2018). The instruments utilized in this research include an expert validation sheet and a metacognition questionnaire. Four instrument validators, comprising experts in the domains of evaluation, development, and experienced lecturers, were involved in the validation process. The instrument trials were administered to a preliminary sample of 10 students on a smaller scale, followed by a broader-scale distribution to 61 students enrolled in research methodology courses. Furthermore, for instrument analysis using Aiken validation on expert validation scores, analysis of the differentiation of instrument items is carried out on the results of small-scale instrument trials to a total of 10 students using Microsoft excel with the formula:

$$d = P_a - P_b = \frac{n_a}{Na} - \frac{n_b}{N_b} \text{..} \text{ (Azwar, 2016:132)}$$

Where:
d = differential power
P_a = Upper group difficulty level
P_b = Lower group difficulty

During the extensive-scale trial, a total of 61 students were administered the metacognition questionnaire, and the data obtained were analyzed using the R programming language. Analysis of instrument item characteristics on a wide-scale trial includes item location parameter coefficients, instrument reliability, proportion of answer selection, ICC per item, Item Information Function (IIF) and total item Standard Error (SE) values.

RESULTS AND DISCUSSION

Instrument form (grid)

The instrument development process commences with the preparation of the instrument framework. The instrument lattice is based on the theory of metacognition and the results and input during *Focus Group Discussion* (FGD) activities with several lecturers who teach research methodology courses. Table 1 below shows the lattice of metacognition skills assessment instrument developed.

Table 1. The lattice of the developed metacognition skills assessment instrument

No.	Component	Sub-component	Item No.	Statement Code
1	Planning	*Determine the goal of the problem* (Determining the purpose of the problem)	1, 2	+ +

		Access background information (Accessing background information on the problem)	3, 4, 5	+ - +
		Allocate resources (Allocate resources that can be used in problem solving)	6, 7, 8	+ + +
		Budget time (Budgeting Time)	9, 10, 11	- - -
2	Monitoring	*Self testing* (Self Testing)	12, 13, 14, 15, 16, 17	+ + + + - +
		Comprehension of task performance (Understanding Task Performance)	18, 19, 20, 21	+ + + +
3	Evaluation	*Appraise Product* Assessing the Problem Solving Results Obtained	22, 23	+ -
		Re-evaluate goals and conclusions (Re-evaluate the problem and make a conclusion)	24, 25	+ +

The developed instrument consists of 25 items and is structured around three primary components of metacognition skills: planning (11 items), monitoring (10 items), and evaluation (4 items). This framework is presented in the form of an instrument that will subsequently undergo validation and review by multiple experts.

Instrument review by experts

The process of reviewing the instrument involved four experts, with each expert specializing in specific domains, including subject matter expertise, linguistic expertise, evaluation expertise, and design expertise. The analysis of the results obtained from this expert validation was conducted using the Aiken validation method. Presented below are the outcomes from the expert validation analysis of the developed instrument.

Table 2. Results of Expert Validation on Developed Instruments through Aiken validation

Validator	**Aspect Score**																			
	Material								**Design**						**Language**					
	M1	**S**	**M2**	**S**	**M3**	**S**	**M4**	**S**	**D1**	**S**	**D2**	**S**	**D3**	**S**	**B1**	**S**	**B2**	**S**	**B3**	**S**
I	5	4	3	2	3	2	3	2	3	2	4	3	5	4	3	2	4	3	4	3
II	3	2	4	3	3	2	3	2	3	2	5	4	5	4	3	2	3	2	4	3
III	4	3	3	2	4	3	5	4	5	4	4	3	5	4	5	4	4	3	4	3
IV	5	4	5	4	4	3	5	4	5	4	5	4	3	2	3	2	3	2	3	2
∑S	13		11		10		12		12		14		14		10		10		11	
V	0,81		0,69		0,63		0,75		0,75		0,88		0,88		0,63		0,63		0,69	
Criteria	Valid		Valid		Valid		Valid		Valid		Valid		Valid		Valid		Valid		Valid	

According to Table 2, the expert validation of the developed instrument is concentrated on three key aspects: material, design, and linguistic components of the instrument. The material aspects include relevance (M1), accuracy (M2), completeness of presentation (M3) and the basic concept of the instrument (M4). Design aspects include

general appearance (D1), special appearance (D2) and presentation (D3). Language aspects include language use (B1), accuracy (B2) and consistency (B3). The validation results showed that all aspects were declared valid and categorised as good by the experts by obtaining Aiken's coefficient of more than 0.60 for all items (Azwar, 2014).

Instrument item parameters

Following the validation process conducted by experts, the instrument underwent a preliminary trial involving a sample size of 10 students. This small-scale trial uses a *paper-based test*. Furthermore, an analysis of the level of difficulty / item location parameters was carried out. The results of the small-scale trial analysis are shown in Table 3 below:

Table 3. Coefficients of grain location parameters at small scale

Grain	Coefficient	Criteria	Grain	Coefficient	Criteria
1	0,3	M	14	0,4	M
2	0,3	M	15	0,4	M
3	0,3	M	16	0,3	M
4	0,3	M	17	0,5	M
5	0,2	TM	18	0,3	M
6	0,3	M	19	0,3	M
7	0,3	M	20	0,3	M
8	0,4	M	21	0,2	TM
9	0,4	M	22	0,2	TM
10	0,4	M	23	0,5	M
11	0,5	M	24	0,3	M
12	0,4	M	25	0,4	M
13	0,7	M			

Description: M = Meets, TM = Does Not Meet

As observed in Table 3 above, among the 25 instrument items that were developed, three items have a differential power coefficient below 0.3, indicating that these items do not meet the required criteria and are considered suboptimal. However, the remaining 22 items meet the criteria and are considered acceptable. However, the researcher still included these items in the broad-scale trial to see how the characteristics of the items when tested on a larger sample.

Subsequently, the instrument underwent an extensive-scale testing phase, involving a larger sample size than in the previous trial. Specifically, the wide-scale trial encompassed 61 students from the History Education, PGSD (Primary School Teacher Education), and Mathematics Education research programs, who participated in completing the instrument. At the analysis stage of the results of the broad-scale trial, item analysis through polytomous IRT (*Library mirt*) and using the *R* software program. Some of the things that were analysed on the results of this broad-scale trial were the reliability of the instrument items, the proportion who chose each answer choice, item location parameters, item fit, and *item information Function* (IIF). The following is a description of each item analysed.

Instrument item reliability

In order to assess the reliability of each instrument item, the internal consistency coefficient, specifically Cronbach's Alpha, was utilized. Items, or all items, with a Cronbach's Alpha coefficient of 0.70 or higher are considered to be reliable and highly dependable (DeVellis, 2016). The reliability coefficients for each item are provided in Table 4 below:

Table 4. Cronbach's Alpha Coefficient on each Item

If	Item	Value	If	Item	Value
Excluding	But_1	0.9431	Excluding	But_14	0.9419
Excluding	But_2	0.9415	Excluding	But_15	0.9429
Excluding	But_3	0.9412	Excluding	But_16	0.9426
Excluding	But_4	0.9452	Excluding	But_17	0.9427
Excluding	But_5	0.9427	Excluding	But_18	0.9423
Excluding	But_6	0.9425	Excluding	But_19	0.9434
Excluding	But_7	0.9432	Excluding	But_20	0.9426
Excluding	But_8	0.9416	Excluding	But_21	0.9424
Excluding	But_9	0.9437	Excluding	But_22	0.9448
Excluding	But_10	0.9461	Excluding	But_23	0.9423
Excluding	But_11	0.9464	Excluding	But_24	0.9430
Excluding	But_12	0.9420	Excluding	But_25	0.9423
Excluding	But_13	0.9447	**All items**		**0.9453**

The data presented in Table 4 indicates that the reliability coefficient for all instrument items collectively is 0.9453, surpassing the standard criteria by a considerable margin and falling well within the category of very high reliability. Furthermore, each individual item contributes significantly to the overall reliability of the instrument, yielding a satisfactory average. When viewed per item, it can also be said that the item that has the lowest contribution to the reliability of the instrument is item 3 (But_3) because if it is removed from the analysis, the reliability of the instrument becomes 0.9412. However, this does not really affect the change in the reliability value of the instrument.

Proportion of answer selection by respondents

This proportion offers insights into the performance of the response options within each instrument item. In this scenario, there are four available choices, denoted as follows: TP = 1, P = 2, S = 3, and SS = 4. A good item is an item where all the answer options are chosen by all respondents or spread across all options. The following are the results of the analysis of the proportion of answer choices by respondents.

Table 5. Proportion of responses to all answer options

But_1	Options	1	2	3	4	But_14	Options	1	2	3	4
	Proportion	0,03	0,34	0,30	0,33		Proportion	0,07	0,28	0,46	0,20
But_ 2	Options	1	2	3	4	But_15	Options	1	2	3	4
	Proportion	0,05	0,33	0,38	0,25		Proportion	0,05	0,33	0,46	0,16
But_ 3	Options	1	2	3	4	But_16	Options	1	2	3	4
	Proportion	0,05	0,34	0,38	0,23		Proportion	0,31	0,43	0,21	0,05
But_ 4	Options	1	2	3	4	But_17	Options	1	2	3	4

	Proportion	0,23	0,38	0,33	0,07		Proportion	0,03	0,41	0,44	0,11
But_ 5	Options	1	2	3	4	But_18	Options	1	2	3	4
	Proportion	0,05	0,33	0,44	0,18		Proportion	0,10	0,43	0,31	0,16
But_6	Options	1	2	3	4	But_19	Options	1	2	3	4
	Proportion	0	0,41	0,41	0,18		Proportion	0,23	0,49	0,20	0,08
But_7	Options	1	2	3	4	But_20	Options	1	2	3	4
	Proportion	0,03	0,51	0,28	0,18		Proportion	0,03	0,51	0,33	0,13
But_8	Options	1	2	3	4	But_21	Options	1	2	3	4
	Proportion	0,05	0,39	0,31	0,25		Proportion	0,13	0,34	0,38	0,15
But_9	Options	1	2	3	4	But_22	Options	1	2	3	4
	Proportion	0,11	0,48	0,34	0,07		Proportion	0,41	0,36	0,18	0,05
But_10	Options	1	2	3	4	But_23	Options	1	2	3	4
	Proportion	0,30	0,38	0,30	0,03		Proportion	0,08	0,26	0,44	0,21
But_11	Options	1	2	3	4	But_24	Options	1	2	3	4
	Proportion	0,31	0,39	0,23	0,07		Proportion	0,07	0,26	0,34	0,33
But_12	Options	1	2	3	4	But_25	Options	1	2	3	4
	Proportion	0,10	0,41	0,31	0,18		Proportion	0,07	0,28	0,38	0,28
But_13	Options	1	2	3	4						
	Proportion	0,18	0,36	0,36	0,10						

According to the data in Table 5, it is apparent that only item 6 (But_6) has one of its response options that was not chosen by the respondents. This suggests that item 6 exhibits a deficiency in terms of response proportion. However, the other 24 items have all the answer options selected by the respondents and it can be said that these 24 items are very good in terms of the proportion of answer selection.

Instrument item location parameters

The subsequent item analysis outcome involves examining the location parameter for each item. In the context of a dichotomous scoring system, this location parameter serves as a discriminative factor. On a polytomous score, an item is said to be good if the location parameter value is greater than 0.25 (Wills, Hambleton, & Purwono, 2018).

Table 6. Parameter coefficients of grain location

Grain	Parameter location	Grain	Parameter location
But_1	-1,22344152	But_14	-0,66938821
But_2	-0,88591727	But_15	-0,67653898
But_3	-0,83388199	But_16	1,14466699
But_4	0,70407859	But_17	-0,61442899
But_5	-0,72458717	But_18	-0,23251314
But_6	0,63623928	But_19	0,69827743
But_7	-0,74729828	But_20	-0,58639686
But_8	-0,82539171	But_21	-0,08709735
But_9	0,35266302	But_22	1,3432613
But_10	1,24532756	But_23	-0,60767677
But_11	0,96143057	But_24	-0,9611059

But_12	-0,29511996	But_25	-0,85427521
But_13	0,3236978		

Based on table 6, it can be explained that the items that have more coefficients than the coefficient. Conversely, the remaining items exhibit coefficients below 0.25, totaling nine items specifically: items 4, 6, 9, 10, 11, 13, 16, 19, and 22. Consequently, in terms of the location parameter analysis, only these nine items can be classified as falling within the "Good" category. The state of each instrument item can also be seen from the following figure 1 which shows how the *Item Characteristic Curve* (ICC) per item / item on the instrument developed.

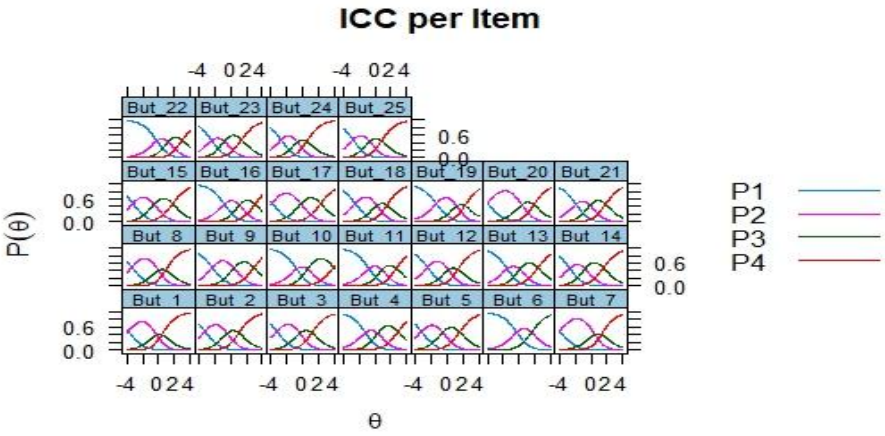

Figure 1. ICC per instrument item

Figure 1 illustrates the varying characteristics of each instrument item. For instance, item 4 (But_4) can be described as having a response pattern that is highly consistent and close to the ideal standard. This means that respondents who have low metacognition skills choose the P1 (Never) option more, then further when their skills are getting to the right up) then in line with the choice of options from the item, namely sequential P2 = Never, P3 = Often, and P4 = very often. This pattern also occurs in items 6, 9, 10, 11, 13, 16, 19, and 22. Furthermore, the attributes of the instrument items can also be observed in the coefficients and the Item Information Function (IIF) diagrams provided below:

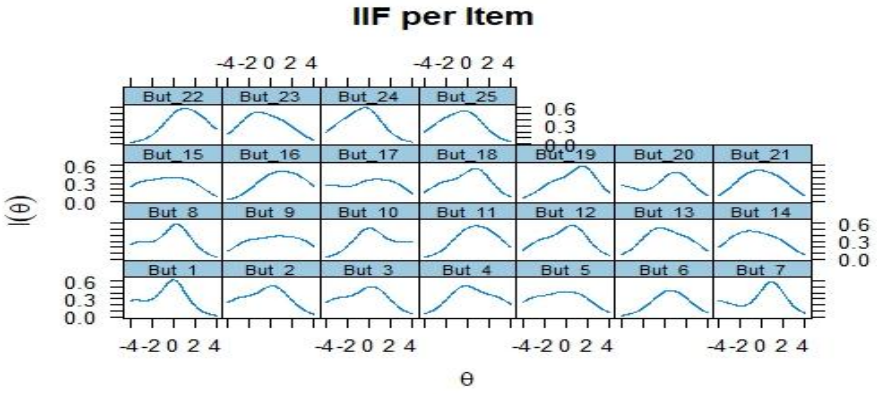

Figure 2. Item Information Function on each instrument item

Figure 2 conveys that each item is capable of furnishing information about the abilities of test takers at various levels. A desirable item is one that exhibits an Item Information Function (IIF) curve resembling a normal distribution curve. From the above results it can be said that there are some items that are less strong in providing information about test takers such as items 5, 9, 15 and 17. While other items are able to provide information that tends to be good.

In general, the instrument items' capacity to offer insights into the abilities of test takers is reflected in Table 7 below:

Table 7. *Test Information function* (TIF)

Lower Bound	Upper Bound	Info	Total Info	Proportion	N Items
-4	4	64,94025	74	0,877	25

The table presented above demonstrates that the 25 items of the instrument collectively possess the capability to offer information about the abilities of test takers, with a proportion of 0.877 (equivalent to 87.7%), covering the abilities of participants within a range of -4 to +4. This percentage falls within the high and commendable category. This means that the instrument

developed has the ability to provide information on test results of more than 85% or Good category.

Furthermore, the *standard error* (SE) curve of all the instrument items developed can also be described as follows:

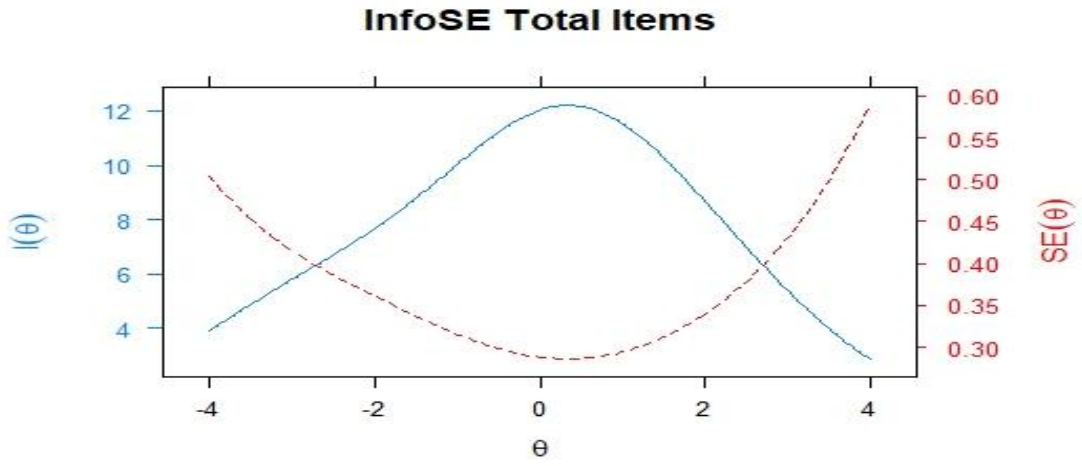

Figure 3. Total *Standard Error* (SE) info

Figure 3 illustrates that the participants' ability (θ) and *standard error* SE (θ) exhibit a high degree of regularity and consistency. In relation to the abilities, the curve resembles a normal distribution, indicating that the number of participants/respondents with abilities both below and above the average is relatively symmetrical. If we look at the SE (θ) curve, the error value ranges from 0.30 to 0.60.

The instrument created for this research is a Likert scale questionnaire consisting of 25 items, each offering four answer choices. This questionnaire was employed to evaluate students' metacognition skills within the context of research methodology courses. The preparation of instruments and items is based on theoretical studies on metacognition, course indicators, and input from lecturers who teach research methodology courses during Focus Group Discussion (FGD) activities. Some of the most urgent input is about the suitability between item statements and course indicators and metacognition itself. In instrument development research, FGD activities are mandatory activities because they aim to get input from various views on the draft and instrument development needs.

The outcomes indicate that the developed instrument meets the criteria for both validity and reliability. The instrument's validity is evidenced by the outcomes of expert validation, where the average Aiken coefficient exceeds 0.60, signifying a high level of validity (S. Azwar, 2014). Additionally, the validity of the individual instrument items is reflected in the distribution of answer choices for each item, with only one item, item number 6, displaying an uneven distribution in answer selections. In addition, the validity of the instrument items is also seen from the item location parameter coefficient. A total of 9 items have a grain location parameter value greater than 0.25 (Wells, et al, 2008) and are also supported by the ICC curve which shows the appropriate pattern on the nine items.

In terms of reliability, all 25 items displayed an Alpha Cronbach reliability coefficient exceeding 0.90. When considering the overall reliability, the instrument achieved an Alpha Cronbach reliability coefficient of 0.94. This reliability coefficient falls within the "Very Good" category (DeVellis, 2016). Furthermore, in the context of the Item Information Function (IIF), each item demonstrated the ability to provide valuable information, with the exception of a few items such as items 5, 9, 15, and 17, which did not exhibit curves closely resembling a normal distribution. Nevertheless, as a collective whole, the instrument items successfully conveyed information from test takers, accounting for more than 85% of the variance.

CONCLUSION

Based on the research outcomes and discussion, this research can conclude that the instrument designed to assess students' metacognition skills in research methodology courses achieved validity, as confirmed by expert evaluations meeting the criteria for goodness. Additionally, it demonstrated validity through the distribution of answer selections across 24 items. Moreover, all the instrument items successfully conveyed information about the conditions of the test participants (respondents), accounting for more than 85% of the variance. While the reliability of the instrument is very high with an Alpha Cronbach score of 0.94.

REFERENCES

Azwar, S. (2016). *Cognitive Ability Test Construction*. Learning Library

DeVellis, R. F. (2016). *Scale development: Theory and applications (Fourth).* SAGE Publications Inc.

Diella, D., & Ardiansyah, R. (2017). The Correlation of Metacognition with Critical Thinking Skills of Grade XI Students on Human Excretion System Concept. *Journal of Science Research and Learning*, 3(2), 134. https://doi.org/10.30870/jppi.v3i2.2576

Hong, W. H., Vadivelu, J., Daniel, E. G. S., & Sim, J. H. (2015). Thinking about thinking: changes in first-year medical students' metacognition and its relation to performance. *Medical Education Online*, 20(1). https://doi.org/10.3402/MEO.V20.27561

Huda, K., Suyanta, Priambodo, Y. A., Ardwiyanti, D., Usman, M., & Alvionita, D. (2021). The Improvement of students' metacognition skills on natural science education using guided inquiry models. *Journal of Physics: Conference Series*, 1882(1). https://doi.org/10.1088/1742-6596/1882/1/012027

Kundre, O. T., & Pratini, H. S. (2019). Students' Thinking Process in View of Their Metacognition Ability in Designing Mathematics Learning. OLIVIA THERESIA KUNDRE, HANIEK SRI PRATINI 1 COMMUNICATION PATTERNS OF SALAFIYAH PESANTREN EDUCATION STUDENT THINKING PROCESS IN TERMS OF M. 1-10.

Mardapi, D. (2018). *Test and Non-Test Instrument Preparation Techniques* (A. Setiawan (Ed.); First). Mitra Cendekia Press.

Mitsea, E., Drigas, A., & Mantas, P. (2021). Soft Skills & Metacognition as Inclusion Amplifiers in the 21st Century. *International Journal of Online and Biomedical Engineering*, 17(4), 121-132. https://doi.org/10.3991/ijoe.v17i04.20567

Ramadhan, A. A., & Pratana, C. (2020). Students' metacognition skill: How the implementation in chemistry learning? *Journal of Physics: Conference Series, 1440(1).* https://doi.org/10.1088/1742-6596/1440/1/012013

S. Azwar. (2014). *Psychological Scale Formulation (Second*). Learning Library.

Wells, C. S., Hambleton, R. K. & Purwono, U. (2008). *Item response theory. Polytomous response IRT models and aplicatios.* Handout delivered on the Training of Educational Assessment and Psychology (Psychometry), at Yogyakarta State University.