JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia), 12(1), 2023, 31–47
DOI: http://dx.doi.org/10.15408/jp3i.v10i1.31679
http://journal.uinjkt.ac.id/index.php/jp3i

Metacognitive Skills Assessment in Research-Proposal Writing (MSARPW) in the Indonesian University Context: Scale Development and Validation Using Multidimensional Item Response Models

Bahrul Hayat[1], Wardani Rahayu[2], Muhammad Dwirifqi Kharisma Putra[3], Iva Sarifah[4], Tina Deviana[5], Valendra Granitha Shandika Puri[1], Khairunesa Isa[6]

Faculty of Psychology, UIN Syarif Hidayatullah Jakarta, Indonesia[1]
Department of Educational Research and Evaluation, Universitas Negeri Jakarta, Indonesia[2]
Faculty of Psychology, Universitas Gadjah Mada, Indonesia[3]
Department of Educational Research and Evaluation, Universitas Negeri Jakarta, Indonesia[4]
Faculty of Psychology, Universitas Mercu Buana Jakarta, Indonesia[5]
Center for General Studies and Co-curricular, Universiti Tun Hussein Onn Malaysia, Malaysia[6]

bahrulhayat@uinjkt.ac.id

Abstract

This study aimed to develop the Metacognitive Skills Assessment in a Research Proposal Writing Context (MSARPW), a multidimensional measure, based on student perspectives, for assessing metacognitive skills in thesis writing, and assess the instrument's psychometric properties using item factor analysis (IFA) and multidimensional item response models. The 40-item MSARPW was administered to 602 Indonesian university students (M_{age} = 25.254 SD_{age} = 6.854). The IFA showed that the two-dimensional factor structure of MSARPW was satisfactory; however, only 24 of the 40 items were found to fit the model. Multidimensional graded response models (MGRM) were applied to the subsequent 24-item MSARPW, which showed that one item (Item 4) did not satisfy the criteria. The estimated reliabilities of each subscale showed that the 23-item MSARPW has good internal consistency (0.891 and 0.902). To conclude, the 23-item MSARPW appears to be a valid and reliable tool for assessing metacognitive skills in the context of research-proposal writing among Indonesian university students.

Keywords: item factor analysis, metacognitive skills, multidimensionality, graded response model, MSARPW questionnaire

Abstrak

Penelitian ini bertujuan untuk mengembangkan Asesmen Keterampilan Metakognitif dalam Konteks Penulisan Proposal Penelitian (MSARPW), suatu ukuran multidimensi berdasarkan perspektif mahasiswa untuk menilai keterampilan metakognitif dalam penulisan tesis, dan menguji karakteristik psikometri instrumen MSARPW dengan menggunakan Item Factor Analysis (IFA) dan multidimensional item response models. Data untuk empat puluh item MSARPW diperoleh dari sampel sebanyak 602 mahasiswa Indonesia (Mean usia = 25.254 SD usia = 6.854). Hasil IFA menunjukkan bahwa struktur faktor dua dimensi dari MSARPW cukup memuaskan; namun, hanya 24 dari 40 item yang cocok dengan model tersebut. Pada tahap berikutnya, Multidimensional graded response models (MGRM) digunakan untuk menganalisis 24 item MSARPW dan hasilnya menunjukkan bahwa satu item (Item 4) tidak memenuhi kriteria. Estimasi reliabilitas untuk tiap subskala menunjukkan bahwa 23 item MSARPW memiliki konsistensi internal yang baik (0,891 dan 0,902). Dapat disimpulkan bahwa 23 item MSARPW merupakan alat ukur yang valid dan reliabel untuk menilai keterampilan metakognitif dalam konteks penulisan proposal penelitian di kalangan mahasiswa Indonesia.

Kata kunci: item factor analysis*, keterampilan metakognitif, multidimensi, model respon bertingkat, kuesioner MSARPW*

Introduction

Metacognition represents awareness of and reflection on one's own cognitive processes, and this behaviour can induce self-regulation and conscious coordination of learning tasks (Flavell, 1979). More specifically, metacognition is a higher-order psychological construct that has been conceptualised as the ability to identify and describe the mental states, beliefs, and intentions of oneself and others (Faustino et al., 2021). In other words, metacognition can be referred to as 'thinking about thinking', or 'cognition about cognitive phenomena', and concerns how an individual knows what they know (Lai, 2011; Roth et al., 2016).

There has been extensive research on metacognition across a variety of research contexts, including different disciplines, age groups, theoretical bases, and cultures (Craig et al., 2020; Lai, 2011; Ohtani & Hisasaka, 2018); the growing development of and interest in this topic can be particularly seen in a recent systematic review of metacognition measures (Craig et al., 2020). From a theoretical perspective, metacognition is generally defined as comprising two intercorrelated aspects: knowledge of cognition (KOC) and regulation of cognition (ROC); the former refer to one's knowledge of their own processing abilities (declarative), ability to solve problems (procedural), and knowledge of when and how to use specific strategies (conditional); where the latter represents the skill to plan learning strategies, manage information, monitor learning, identify mistakes, and evaluate learning (Flavell, 1987; Pintrich, 2004). Although some scholars have forwarded alternative factor structures that suggest that there are more than two aspects of metacognition (e.g. Desoete et al., 2001), these two abovementioned correlated factors of metacognition are commonly used in the development of associated measurement instruments (e.g. Teng, 2020).

Regarding samples, previous studies have generally been conducted in specific contexts and among specific samples, such as university students (e.g. Teng, 2020) or clinical samples (e.g. Papegeorgiou & Wells, 2003). Notably, one recent study revealed that students lack metacognitive knowledge (Anthonysamy et al., 2021); this is despite the fact that metacognitive skills have been shown to be strongly related to academic performance (Craig et al., 2020), which is, in turn, closely related to intelligence (Ohtani & Hisasaka, 2018; Song et al., 2021). Thus, there is strong evidence that metacognition is important in the context of education.

In the context of higher education, metacognition is a tool that not only helps students involve themselves in the learning process, but also provides them with autonomy regarding learning tasks (Roth et al., 2016). Metacognitive skills have been found to improve learning by helping students advance their understanding and achieve better academic outcomes (Firdayani et al., 2020; Wolters & Hussain, 2015). In particular, students with strong metacognitive abilities can, through applying appropriate strategies, manage, plan, and structure their learning through thought management, learning assessment, and calculation of the time needed to learn (Roth et al., 2016).

For university students, one of the important roles of metacognitive skills concerns scientific writing (Teng, 2020); there are various forms of scientific writing, including theses, dissertations, and scientific articles (Inouye & McAlpine, 2022). In particular, one of the most important scientific-writing skills for students is the ability to write research proposals, a process that is useful for both teaching and learning research methods (Saeed et al., 2021). Metacognitive skills have also been found to help students complete their final projects (e.g. theses or dissertations), playing a role from the proposal-writing stage to completion (Marhaban et al., 2021). In particular, previous studies have found that the two previously mentioned aspects of metacognition, knowledge of cognition and regulation of cognition, have positive impacts on improving students' writing skills (Cer, 2019).

Students in Indonesia often encounter problems writing final project proposals (Daniel & Taneo, 2019; Kocimaheni et al., 2020; Tutpai & Unja, 2022). This problem arises despite the fact that, in the

Indonesian curriculum, at the undergraduate, graduate, and doctoral levels research proposal seminar courses are mandatory subjects offered one semester before the final semester (see Faculty of Psychology University of Indonesia, 2020). Students are required to pass the research proposal seminar course with large credits (e.g. 6 credits) before taking the final course (e.g. dissertation with 18 credits). If a student does not pass the course, his/her study time will increase, delaying the completion of their studies (e.g. van de Schoot et al., 2013; van Rooij et al., 2021).

As stated by Inouye and McAlpine (2022), academic writing, such as research-proposal writing, requires extensive analytical, research, and communication skills, and multiple studies have reported on the challenges associated with learning such skills (e.g. Kuiken & Vedder, 2021; Quvanch & Na, 2022). The thesis, a major project, is created over several years and under the supervision of professors; notably, supervisors can vary in their approaches to writing support, meaning students' academic writing skills are often influenced by the different forms of feedback they receive regarding their writing (Cayley, 2020; Inouye & McAlpine, 2022). Metacognitive skills may help students in such circumstances, as these skills have been proven to help students perform academic writing (e.g. Teng, 2020).

However, it should be noted that, although a systematic review has shown that there are many instruments for measuring metacognition (Craig et al., 2020), such measurement remains challenging because metacognition is not an explicit behaviour (Teng, 2020). For example, several instruments have been developed over the past three years to measure metacognition, metacognitive awareness, and metacognitive skills; these include the Metacognitive Awareness Scale—Domain Specific (MCAS-DS; Song et al., 2021), the Metacognitive Skills Inventory (MSI; Hameed & Cheruvalath, 2021), the Metacognition Self-Assessment Scale (MSAS; Faustino et al., 2021), and the Metacognitive Writing Strategies Questionnaire (MWSQ; Teng, 2020). The MSAS and MSI are developed for general contexts, while the MCAS-DS and MWSQ relate to certain specific contexts. The present authors consider the latter instrument to provide a suitable instrument-development framework for adaptation to the context of research-writing proposals.

The MWSQ was developed with the involvement of students, which helped to clarify their general strengths and weaknesses in regard to academic writing (Teng, 2020). While previous studies have found that, in writing thesis proposals, many students encounter difficulties determining research topics, choosing research designs, classifying the thesis-proposal genre/type, and evaluating literature (Xia & Luxin, 2012), we chose to involve students in our item-generation processes in order to identify the kinds of difficulties students in Indonesian universities face when writing proposals. Thus, the difficulties encountered by such students was obtained directly, and not based on the findings of previous studies, which could have limited the accuracy of our tool.

In terms of methodological perspectives, to obtain additional information about the characteristics of items and samples, not only for unidimensional measures, but also multidimensional measures, current psychometric practice encourages the use of item-response models (e.g. Immekus et al., 2019; Yau et al., 2015). As, theoretically, metacognition comprises two related components (e.g. Flavell, 1987), multidimensional item-response models are suitable for use in this regard. This methodology, when compared to descriptive information alone, can provide greater information on the relationship between an item and latent factors (Hayat et al., 2021). A recent study on measurement of self-regulated learning in students, in which metacognition was one aspect, implemented item-response theory (IRT) in analysing the data, and provided more detailed analysis than that produced through the classical approach (e.g. classical test theory; Uesaka et al., 2022). However, to the best of our knowledge, no previous studies have applied IRT in the process of validating metacognition measurement instruments in the context of research-proposal writing. Thus, the present study will be the first to measure such a contribution of IRT models.

Thus, the purpose of the present study was to develop an instrument that can measure metacognitive skills relating to research-proposal writing among university students, and to validate the instrument using multidimensional item-response models. A multidimensional graded response model was used to investigate the overall model fit, item parameters, item fit, reliabilities, and item characteristic curves. We believe that the implementation of such a combined approach will result in the creation of a stronger instrument that can capture metacognitive skills for very specific situations and sample characteristics.

Metacognitive Skills Assessment in Research Proposal Writing Context: Scale Development Processes

The Metacognitive Skills Assessment in Research Proposal Writing Context (MSARPW) questionnaire, which was developed by the present authors, was used to measure metacognitive knowledge in the context of research-proposal writing. Several steps were taken to develop the questionnaire. The items for the MSARPW were initially developed through applying a combination of techniques. This combination of techniques included the involvement of student participants, who were asked several questions related to metacognitive knowledge, such as their strengths and weaknesses regarding writing thesis proposals and the difficulties they faced in terms of proposal writing. A total of 32 university students taking courses in 'research proposal writing' (*seminar proposal penelitian*) participated in this process. After completing some descriptive questions, findings from these participants were presented to six lecturers, who proceeded to identify the key information concerning the item-writing process, and to categorise the findings regarding difficulties into nine themes. Descriptive findings regarding the difficulties the students experienced writing research proposals are presented in Table 1.

Table 1. Descriptive Data for the Types of Difficulties the Students Experienced Writing Research Proposals.

No	Difficulty Type
1	Writing appropriate reasons for choosing a specific research design
2	Recalling relevant details from courses taken before writing the proposal
3	Maintaining consistency in writing across the research problem, research aims, research question, and hypotheses
4	Using reference-management software, which is often mandatory for universities
5	Evaluating whether the data-analysis method used is appropriate for answering the research question
6	Knowing whether the literature review has sufficient depth
7	Formatting the research report to the standards required by the university
8	Meeting the writing standards of APA 7th edition
9	Rephrasing and summarizing theoretical reviews using one's own words

We combined the nine themes shown in Table 1 with the two-aspect framework of metacognition (knowledge of cognition and regulation of cognition; e.g. Teng, 2020). Knowledge of cognition comprises three skills: knowledge of one's own processing abilities (declarative), the ability to solve problems (procedural), and knowledge of when and how to use specific strategies (conditional). Meanwhile, regulation of cognition includes planning, monitoring, and evaluating skills (Craig et al., 2020; Teng, 2020).

Based on the item-writing strategies obtained from combining the nine difficulty themes with the three skills contained within each subscale, 51 items related to metacognition in research-proposal writing were created. All items were written and administered in the Indonesian language. Lastly, the items were evaluated for clarity and readability by an expert in psychological measurement. This expert deemed 11 items to be ambiguous and/or out of context and, thus, we eliminated these, leaving 40 items. Thus, the final MSARPW instrument contained 40 items, all of which were scored using a four-point Likert scale (1 = 'strongly disagree', 4 = 'strongly agree'), and we investigated the behavioral manifestation of each aspect.

Methods

Participants

Data were gathered from 602 university students (311 males [51.7%], 291 females [48.3]), who were recruited from 10 universities across Indonesia. The mean age of the sample was 25.254 years (SD = 6.854). All of the respondents were active students in their universities' faculties of education, with 476 (79.1%) being undergraduate students, 81 (13.5%) being graduate students, and 45 (7.5%) being doctoral students. The data were collected using an online approach. Students were given information regarding the general aim of the study, and were assured that the data would be handled in a manner that protected their privacy. All students participated on a voluntary basis, and no incentive or compensation was provided to them for their participation. Also, informed consent was obtained from all individual participants included in the study.

Data Analysis Methodology

Item Factor Analysis

The first analysis was a validity check of the internal structure of the MSARPW, which was performed using item factor analysis (IFA), which has long been used in educational and psychological research to explore the theoretical dimensions of measurement instruments (Wirth & Edwards, 2007). IFA comprises factor analysis of categorical item-level data, and features both exploratory and confirmatory approaches (Cai, 2010; Rifenbark et al., 2021). In confirmatory IFA, a specific hypothesized factor structure is proposed (including the correlations among the factors) and then statistically evaluated. If the estimated model fits the data, then a researcher concludes that the factor structure replicates (Reise et al., 2000).

In this study, we used several statistics and fit indices: root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker–Lewis index (TLI), and standardised root-mean-square residual (SRMR). Based on previously published criteria (Marsh et al., 2005), the following standards for good fit were set: CFI > 0.95, TLI > 0.95, RMSEA < 0.06, and SRMR < 0.08. To perform our factor analysis, we proposed two-correlated factor model as a hypothesised model. IFA was also used to estimate item and person parameters; this was implemented using Mplus 8.3. When using Mplus, parameters were estimated using weighted least squares means and variance adjusted (WLSMV), which was specifically developed for latent variable models with ordinal categorical responses (see Li, 2016).

Multidimensional Graded Response Model

In the early years of its development, a key assumption in applying item response modelling was construct unidimensionality (e.g. Bock & Gibbons, 2021; Samejima, 1969). Generally, testing unidimensionality assumptions in an item response model is performed using the goodness-of-fit index specifically developed for IRT and combining this with statistics that are commonly used in factor analysis (e.g. Kircaburun et al., 2021), exploratory and confirmatory factor analysis methods (e.g. Hayat et al., 2021; Suseno et al., 2022), or other statistics for dimensionality testing that have been specifically developed for IRT models (e.g. Chou & Wang, 2010; de Ayala, 2022).

However, as the MSARPW has a multidimensional structure with a large number of items, commonly used unidimensional versions of item response models, such as the graded response model (GRM; Samejima, 1969) or partial credit model (Masters, 1982), cannot be used because it would be difficult for such models to fulfil the unidimensionality assumptions associated with the multidimensionality nature of MSARPW.

Fortunately, a multidimensional extension of the GRM (MGRM; de Ayala, 1994) has been developed. The MGRM requires a set of multidimensional parameters. In the MGRM, examinee responses to item i are categorised into $m_i + 1$ ordered categories, in which higher categories indicate greater θ level, and m_i is the number of category boundaries. The category scores for item i, x_i take the values 0, ..., m_i. The MGRM is expressed as follows:

$$P_{x_i}(\Theta) = \frac{\exp[D\Sigma a_{ih}(\theta_h + d_{xi})]}{1 + \exp[D\Sigma a_{ih}(\theta_h + d_{xi})]}$$

where θ_h is the latent trait on dimension h (h = 1, ..., *r* dimensions), a_{ih} is the discrimination parameter for item i on dimension h, d_{xi} is the difficulty parameter for category x for item i, and the summation is across dimensions (de Ayala, 1994). In the context of MSARPW, three threshold parameters are generated for each item. The MGRM model also shares assumptions of local independence and monotonicity, as in the unidimensional GRM model.

According to Standard 3.9 of the Standards for Educational and Psychological Testing (AERA et al., 2014), both the overall model fit index as well as the item level should be reported when an IRT model is used. In the present study, the model fit indices used for the entire model were M_2 and $RMSEA_2$. If M_2 was found to be non-significant, this would indicate that the model fit the data (Maydeu-Olivares & Joe, 2006); meanwhile, if $RMSEA_2$ showed a value of < 0.05, this would also indicate that the MGRM model fit the data (Maydeu-Olivares, 2013). Reliability in IRT is described by the coefficient of marginal reliability (Green et al., 1984), which is analogous to the alpha coefficient in classical test theory (Reise, 1999); if the value is higher than 0.80, the instrument has good internal consistency (e.g. Petscher et al., 2015). After obtaining the fit model for the entire model, the fit model was tested at the item level using the $S-\chi^2$ method (Kang & Chen, 2008); this method has been tested for its performance and has been proven to perform well in MGRM (Su et al., 2021). Items were fit to the model if $S-\chi^2$ was not significant at $p < 0.05$ (Kang & Chen, 2008), and other opinions were $p > 0.01$ (Stover et al., 2019).

To estimate item and person parameters, an MGRM was used in this study through using the '*mirt*' (Chalmers, 2012) package in Rstudio. In using the *mirt* package, person and item parameters were estimated using marginal maximum likelihood, which was developed from the IRT approach over a long period (see Bock & Aitkin, 1981).

Results and Discussion

Factor Analysis: Initial Analysis for Item Selection

In the first stage, an initial analysis using limited-information IFA (using the WLSMV estimator) was conducted on the 40-item MSARPW to determine whether the items compiled measured the intended variables. The results of this analysis using the two-correlated factor model revealed that the model did not fit the data, showing poor goodness of fit (χ^2_{WLSMV} [739] = 3738.055, $p < 0.000$; RMSEA = 0.082, 90% CI [0.080, 0.085]; CFI = 0.770; TLI = 0.757; SRMR = 0.088). We found that several items had low and insignificant factor loadings, while others showed negative factor loadings. Based on these IFA results, 16 of the 40 items were removed due to non-significant and/or negative factor loadings.

Factor Structure: Dimensionality Testing of the Item Response Models

The result of the IFA confirmed the two-correlated factor model (KOC: knowledge of cognition; ROC: recognition of cognition), according with the hypothesised factor structure for the scale, because the values of the indices were above the acceptable threshold (χ^2_{WLSMV} [251] = 495.992, $p < 0.000$; RMSEA = 0.040, 90% CI [0.035, 0.045]), CFI = 0.974, TLI = 0.971, SRMR = 0.038). Thus, based on the RMSEA, CFI, TLI and SRMR values, the results indicated that the two-correlated factor model was satisfactory and representative of the underlying structure of the MSARPW. All items loaded significantly (ranging from 0.505 to 0.805) in relation to both latent factors at a $p < 0.01$ significance level (see Figure 1).

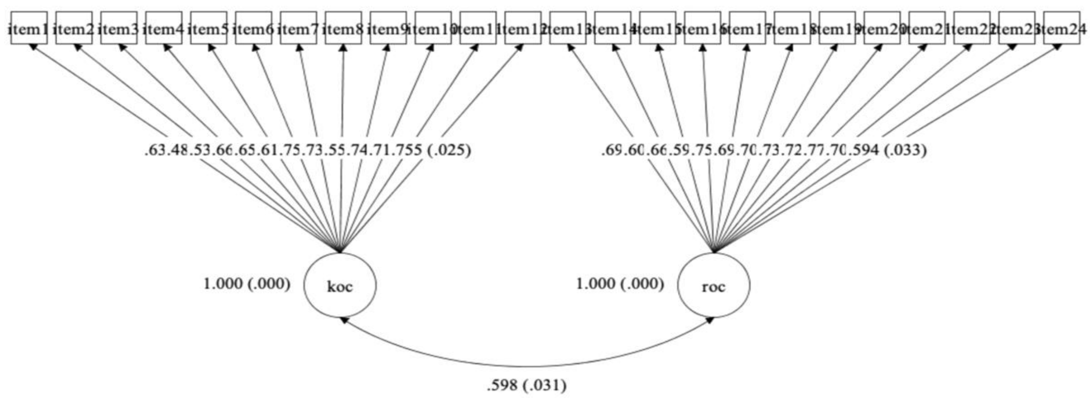

Figure 1. Two-Correlated Factor IFA Model.

MGRM Analysis Results

Item Parameters

Based on multidimensional factor structure evidence obtained through the IFA, calibration of the MSARPW items was performed using multidimensional item response models; specifically, MGRM. The result of the MGRM analysis confirmed the two-factor model, because the values of the indices exceeded the acceptable threshold (M_2 [203] = 465.315, $p < 0.000$; $RMSEA_2 = 0.047$, 90% CI [0.042, 0.053], CFI = 0.968, TLI = 0.964). This accorded with the hypothesised factor structure of the scale and the IFA findings. Also, based on IFA we found no residual correlation added to the models, indicating that the local independence assumption was fulfilled. Table 2 shows the results of the analysis of the MSARPW, including item discrimination and threshold parameters.

Table 2. Item Wording and Estimated Parameters of the MSARPW.

No	Original item numbers	a1	a2	d1	d2	d3	b1	b2	b3
1	ITEM 1	1.503	0.000	1.125	0.056	−1.634	-0.749	-0.037	1.088
2	ITEM 2	1.009	0.000	1.939	−0.352	−1.358	-1.921	0.349	1.346
3	ITEM 3	1.122	0.000	0.798	−0.441	−1.944	-0.711	0.393	1.733
4	ITEM 4	1.583	0.000	2.023	−0.782	−2.086	-1.278	0.494	1.318
5	ITEM 5	1.546	0.000	1.590	−0.416	−3.073	-1.029	0.269	1.988
6	ITEM 6	1.427	0.000	2.095	−0.960	−3.243	-1.469	0.673	2.273
7	ITEM 7	1.874	0.000	2.700	−0.064	−1.787	-1.441	0.034	0.953
8	ITEM 8	1.696	0.000	1.615	−0.384	−1.122	-0.952	0.226	0.661
9	ITEM 9	1.153	0.000	2.223	−0.300	−1.916	-1.928	0.260	1.661
10	ITEM 11	1.998	0.000	2.040	−0.027	−1.899	-1.021	0.013	0.951
11	ITEM 13	1.882	0.000	1.784	−0.551	−3.359	-0.948	0.293	1.785
12	ITEM 14	2.090	0.000	1.359	−0.630	−2.130	-0.650	0.302	1.019
13	ITEM 16	0.000	1.614	2.505	0.206	−1.811	-1.552	-0.128	1.122
14	ITEM 17	0.000	1.383	0.828	−0.429	−1.877	-0.599	0.310	1.357
15	ITEM 20	0.000	1.662	2.416	−0.712	−3.074	-1.453	0.428	1.849
16	ITEM 22	0.000	1.360	1.714	−0.838	−2.718	-1.261	0.616	1.999
17	ITEM 24	0.000	1.981	1.770	−0.690	−2.673	-0.894	0.348	1.349
18	ITEM 25	0.000	1.688	1.389	−0.830	−2.874	-0.823	0.492	1.702
19	ITEM 26	0.000	1.678	2.467	0.352	−2.143	-1.470	-0.210	1.277
20	ITEM 35	0.000	1.724	1.828	0.011	−1.944	-1.061	-0.006	1.128
21	ITEM 36	0.000	1.933	1.812	−0.180	−2.244	-0.937	0.093	1.161
22	ITEM 38	0.000	2.096	1.883	−0.418	−2.477	-0.898	0.199	1.181
23	ITEM 39	0.000	1.830	1.682	−0.123	−3.175	-0.919	0.067	1.735
24	ITEM 40	0.000	1.350	1.491	−0.461	−2.264	-1.105	0.341	1.678

Table 2 contains information regarding the item discrimination parameters (a1 and a2) and thresholds (d1, d2, and d3 or b1, b2, and b3 in IRT parameterization) for each response option for each item. The results of the MGRM analysis showed that none of the items had a negative discriminatory power, indicating that the items could function well in regard to distinguishing people with high metacognition from those with low metacognition. In addition, the discrimination items were in an ideal range (a = 1.009–2.096). Item 38 (a = 2.096; 'Saya melakukan evaluasi apakah pertanyaan penelitian sudah sejalan dengan latar belakang masalah, tujuan penelitian dan hipotesis' ['I evaluate whether the research questions accord with the background of the problem, research objectives, and hypotheses']) and Item 24 (a = 1.981; 'Saya memiliki strategi yang terencana untuk memilih metode penelitian yang sesuai untuk menjawab pertanyaan penelitian yang dirumuskan' ['I have a well-planned strategy for selecting appropriate research methods for answering the research questions that have been formulated']) were the two items most sensitive for distinguishing people with low regulation of cognition and people with high regulation of cognition. Meanwhile, Item 14 (a = 2.090; 'Saya memiliki strategi yang efektif untuk menyeleraskan judul penelitian, rumusan masalah, tujuan penelitian dan hipotesis yang akan uji' ['I know an effective strategy for aligning the research title, statement of the problem, research objectives and hypotheses to be tested']) and Item 11 (a = 1.998; 'Saya mengetahui bagaimanakah cara penulisan latar belakang masalah yang sistematis agar mudah diselaraskan dengan formulasi tujuan penelitian' ['I know how to write background of the problem in a systematic way so that it is easy to formulate research objectives']) were the two items most sensitive for distinguishing people with low knowledge of cognition from those with high levels of cognition. We found that the four abovementioned items measured the same form of difficulties regarding aligning research problems, aims, questions, and hypotheses.

Conversely, Item 2 (a1 = 1.009; 'Saya tahu bahwa saya mampu untuk menulis proposal penelitian dengan menggunakan standar penulisan (e.g. panduan penulisan yang disediakan kampus and APA 7th edition)' ['I know that I am capable of writing a research proposal that meets accepted writing standards (e.g. the writing guidelines provided by the campus and APA 7th edition)'] and Item 40 (a2 = 1.350; 'Saya selalu memantau perkembangan terkini tata cara penulisan ilmiah agar selalu menerapkan standar terbaru dalam penulisan proposal ilmiah' ['I always monitor developments in scientific writing procedures so that I can apply the latest standards in writing scientific proposals']) were the items with the lowest discriminatory power. This means that these items related to behaviours that were not very sensitive in regard to distinguishing people with low and high regulation of cognition or knowledge of cognition. However, it should be noted that the item discrimination values of Item 2 (a1 = 1.009) and Item 40 (a2 = 1.350) were in the acceptable criteria, indicating well-functioned items.

Furthermore, threshold values are presented below (in Table 2), ordered from lower to higher, and were valid for all items. This pattern indicates that the monotonicity assumption in this study was fulfilled. This assumption would not have been fulfilled if any category had not been selected by at least one respondent. A graphical representation of the different powers and thresholds for each item is presented in Figure 2.

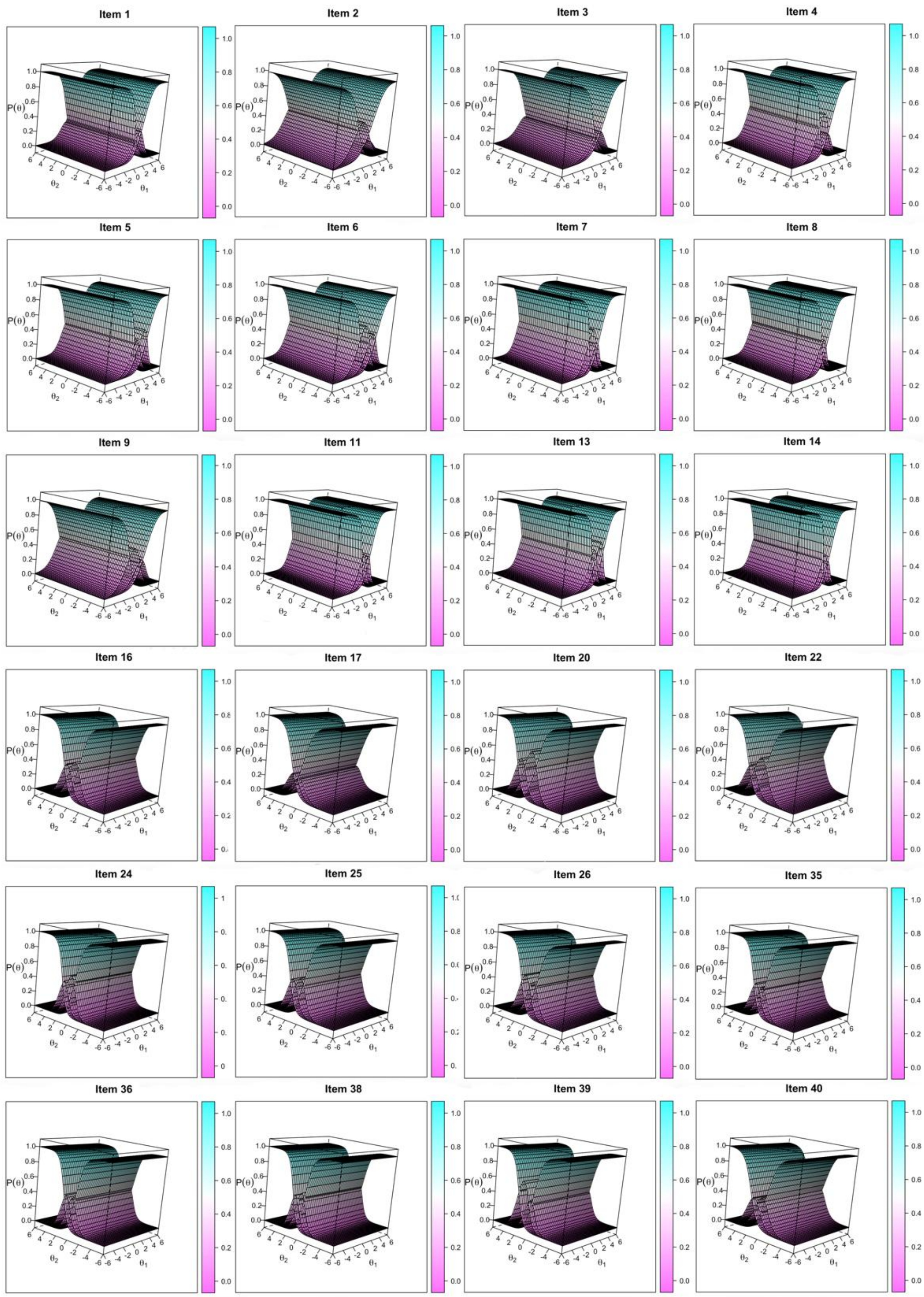

Figure 2. Item Characteristic Curve of MSARPW 24-Items.

As can be seen in Figure 2, the item characteristic curve (ICC) for the multidimensional IRT differed from the ICC for the unidimensional IRT. In the first 13 items, which measured knowledge of cognition, the ICC for each item approached θ1 (the level of knowledge of cognition), while the final 13 items, which measured recognition of cognition, approached θ2 (the level of regulation of cognition). Given that all items had optimal discriminating power, no ICC was inverted or too flat. These findings, which can be considered optimal, were achieved because items with negative factor loadings, which were considered problematic items, were removed in IFA. Thus, given these optimal findings, it can be interpreted that respondents with low trait levels tended to choose response category 1, and respondents with high trait levels tended to choose response category 4. It can be seen in Figure 2 that each category for each item had different range thresholds. For example, there were items for which it was relatively easy to obtain a score of 2 when compared to other items, for which it was more difficult to get a score of 2.

Item Fit Statistics

Based on multidimensional factor structure evidence obtained through IFA, calibration of the MSARPW items was performed using a MGRM. Table 3 shows the results of the calibration of the MSARPW, including MGRM fit statistics such as $S - \chi^2$, with its degrees of freedom and p-value, and also RMSEA $S - \chi^2$. Although all models showed a good fit to the MGRM, Table 3 shows that one item (item 4) had a poor fit to the MGRM when viewed based on the $S - \chi^2$ statistical test. However, all items showed acceptable RMSEA $S - \chi^2$. Therefore, we excluded item 4 based on the findings of the MGRM analysis, resulting in the final 23-item version of the MSARPW.

Table 3. Item Fit Statistics for the MSARPW.

Item	Original item numbers	$S - \chi^2$	df $S - \chi^2$	RMSEA $S - \chi^2$	p-value $S - \chi^2$
1	ITEM 1	114.761	95	0.019	0.082
2	ITEM 2	81.757	105	0.000	0.955
3	ITEM 3	61.246	69	0.000	0.735
4	ITEM 4	89.587	56	0.032	0.003
5	ITEM 5	48.415	46	0.010	0.376
6	ITEM 6	92.574	85	0.013	0.269
7	ITEM 7	43.463	51	0.000	0.764
8	ITEM 8	88.810	89	0.000	0.486
9	ITEM 9	55.672	57	0.000	0.525
10	ITEM 11	64.731	50	0.023	0.079
11	ITEM 13	104.397	85	0.020	0.075
12	ITEM 14	75.744	64	0.018	0.149
13	ITEM 16	56.716	50	0.015	0.239
14	ITEM 17	70.663	66	0.011	0.325
15	ITEM 20	99.427	89	0.014	0.211
16	ITEM 22	68.767	53	0.023	0.071
17	ITEM 24	65.754	53	0.021	0.112
18	ITEM 25	71.657	54	0.024	0.054
19	ITEM 26	40.400	51	0.000	0.857
20	ITEM 35	53.914	54	0.000	0.478
21	ITEM 36	53.473	54	0.000	0.495
22	ITEM 38	88.581	89	0.000	0.493
23	ITEM 39	53.329	40	0.024	0.077
24	ITEM 40	85.780	66	0.023	0.051

Factor Correlations: MGRM and IFA Results

As shown in the lower row of Table 4, the direct estimate for the correlation matrix of the 2-subscale MSARPW, conducted using the multidimensional item response model approach, was 0.607. In general, the direction of the correlations was consistent with the theory of metacognitive skills, where both aspects were positively correlated with each other; this supported the expected outcome regarding the MSRPW's internal structure. The correlation was statistically significant.

Table 4. Latent Correlation between the MSARPW's Subscales: MGRM and IFA.

Variables	KOC	ROC
KOC		0.598
ROC	0.607	

As shown in the upper row of Table 4, the direct estimate of the correlation matrix for the two-subscale MSARPW, conducted using the item factor analysis approach, was 0.598. In general, the pattern of the correlations was consistent with the theory of two-factor metacognition. We suspect that small differences in the correlation derived from MGRM and IFA occurred due to the estimator differences (MML and WLSMV), and conclude that the differences were not substantial.

The present study reported on multidimensional IRT analysis that was conducted to perform psychometric validation of the MSARPW, a new instrument for measuring metacognitive skills in research-proposal writing in the Indonesian university context. In this study, we illustrated how MGRM analysis was applied to analyse the MSARPW, as this instrument featured a multi-dimensional factor structure. The findings from our study indicate that the 23-item MSARPW has excellent psychometric properties.

The final 23-item MSARPW was developed through a relatively long process. Initially, 51 items were created based on analysis of two aspects of metacognition combined with nine themes of difficulties students experience writing proposals. The items were then subjected to expert review, after which only 40 items remained. In the initial analysis using IFA, 16 of these 40 items showed negative or very low factor loadings (e.g. < 0.200), and we consequently decided to remove these 16 items; thus, 24 items remained at this point. The IFA-based item-exclusion approach we applied has also been used in previous studies (e.g. Jans-Beken & Wong, 2021; Mya et al., 2021), meaning our approach is methodologically justified.

Prior to performing MGRM, we again performed IFA to confirm the factor structure of the 24-item MSARPW. This IFA consequently confirmed the multidimensionality of the MSARPW, which supported the use of MGRM. This finding indicated that the resulting metacognitive factor structure accorded with previous studies' observations of two-factor models for metacognition (e.g. Craig et al., 2020; Teng, 2020). However, this finding did not accord with the findings of other studies that observed a third factor; namely, metacognitive beliefs (e.g. Desoete et al., 2021). This difference certainly opens opportunities for future research to examine the concurrent validity of the MSARPW through assessing the relationship between the MSARPW and other metacognitive self-assessment instruments that feature different dimensions or factor structures (e.g. MSI and MSAS).

Through the MGRM analysis, we found excellent model fit indices that accorded with the IFA two-correlated factor structure model. However, one item (item 4) had unacceptable fit indices, meaning 23 items fit the MGRM. All items showed optimal item discrimination, with the range being 1.009–2.096; this satisfied the recommended criteria for ideal item discrimination (< 2.500; Baker & Kim, 2004). We also found that no items had extremely high discrimination (> 4.00), which would have indicated a problem with those items (Edelen & Reeve, 2007). Further, based on reviewing the wording of the items, we realised that the item with the highest discrimination was item 4, which measured difficulty maintaining consistency in writing the descriptions of the research problem, research aims, research question, and hypotheses.

Through a subsequent review, we found that the problem of maintaining consistency across written descriptions of a study's research problem, research aims, research question, and hypotheses has been the subject of a substantial amount of research (e.g. Akhidime, 2017; Dhir & Gupta, 2021; Farrugia et al., 2010; Newman & Covrig, 2013). In writing research, such consistency (particularly across the title, problem, purpose, and research question) is crucial, as it will improve the logic and transparency of the research (Newman & Covrig, 2013). Some previous studies have provided tips for aligning the research question, hypotheses, and objectives (see Dhir & Gupta, 2021; Farrugia et al., 2010). Based on the empirical evidence obtained from the MSARPW items relating to this form of difficulty, we found that this form of difficulty has the highest sensitivity for distinguishing people with low and high regulation of cognition and knowledge of cognition, respectively. This is a novel finding of the present study, and may provide be of use for future diagnosis tools focussing on university students' difficulties writing research proposals.

Regarding reliability, the MSARPW also demonstrated adequate internal consistency for both subscales (expected a posteriori [EAP] reliability = 0.891 and 0.902, respectively). These findings are consistent with those of previous studies that used different instruments that comprised the same two subscales as the MSARPW. Specifically, the MSAS was found to have a Cronbach's alpha range of 0.72–9.87 (Pedone et al., 2017), while the MSWQ was found to have an alpha range of 0.71–0.81 (Teng, 2020). Based on EAP reliability, the MSARPW has high reliability, and we consequently conclude that the MSARPW has excellent internal consistency. However, it should be noted that this reliability does not refer to the measurement instrument itself, but to the consistency of the results obtained (Thompson & Vacha-Haase, 2000); thus, acceptable reliability is present only for our study sample.

Further, regarding the item response models, this study provides an overview of the application of the item response models when the data are multidimensional. Some studies that have used GRM have calibrated separately for each dimension (e.g. Yau et al., 2015), but in the present study we used the MGRM and found important information in the form of correlations between latent variables and no correlation between observed scores. In this approach, correlation between two aspects of metacognition is already corrected for attenuation (Adams et al., 1997). Thus, the use of the MGRM in this study provided an advantage regarding the validation of MSARPW in terms of its internal structure.

In addition, MGRM analyses supported the correlation pattern of the MSARPW factor structures, and we found that the two dimensions of the MSARPW were significantly positively intercorrelated. This finding accords with those of previous research (see Craig et al., 2020), which also reported that both dimensions are positively correlated. In other words, we found that, among Indonesian university students, higher knowledge of cognition is correlated with higher regulation of cognition. The correlation between factors was found to be 0.607, which was higher than the findings of a previous meta-analysis of the correlation between these two factors (0.340; Craig et al., 2020). However, the present study's finding was lower than that obtained by the MWSQ, which found a correlation of 0.800 (Teng, 2020). Nevertheless, the present findings of a correlation between these factors can be utilised in future meta-analyses focussing on this correlation.

The present study's findings have implications for the science of metacognition. If studies use scales that do not specifically assess the construct of metacognition in the context of research-proposal writing (including difficulties writing research proposals according to student perspectives), the outcomes may not show the full picture regarding metacognition as a skill that has an important supportive role in the writing of research proposals. This might be the reason the association between metacognition and research-proposal writing has, to date, not been as clear as the association between self-regulation and research-proposal writing (Arianto & Wulyani, 2022; Mbato & Cendra, 2019) and the association among metacognition and intelligence and academic performance (Ohtani & Hisasaka, 2018). Also, future research should examine the MSARPW in terms of its association with other variables that have been found to be predictors of metacognition in order to determine whether this instrument indeed performs better in regard to identifying research complexities such as causality.

Finally, there are several limitations to the present study that should be taken into consideration when interpreting the results. The study sample comprised university students from 10 Indonesian universities, but was limited to students from faculties of education. Therefore, the present findings should be replicated using different samples from different countries, more diverse universities, and more diverse faculties, schools, and/or age groups. Second, the present study collected data using a self-report online survey, which is susceptible to specific biases (e.g. social desirability and aberrant response). Therefore, future studies should adopt more in-depth methods, such as interviews, to better investigate the difficulties of research-proposal writing. Third, the present study was a cross-sectional validation, which restricted the making of any causal assumptions regarding other variables that may be related to metacognition. Therefore, future studies should examine the relationships among the variables that are hypothesised to impact metacognition of research-proposal writing.

Despite these limitations, however, the present study is the first to develop a psychometrically valid and reliable tool for assessing individuals who experience difficulty writing research proposals, and can be used to assess, among Indonesian university students, metacognitive skills relating to writing research proposals.

Conclusion

In conclusion, the current study confirmed the construct validity of the two-subscale MSARPW. The 23-item MSARPW, in its current state, represents an appropriate scale for providing diagnostic information for university students regarding their metacognitive knowledge and metacognitive regulation capacity in writing research proposals. Academic supervisors may use the information to determine students' metacognitive capacities and consequently provide these students with proper assistance based on their metacognitive capacities. For university students who wish to achieve a certain outcome by learning independently, having a better understanding of their own metacognitive capacity can help them monitor and evaluate their learning progress. It is also hoped that the present study can contribute to improving current understanding of the role metacognition plays in research-proposal writing in the Indonesian university context.

Conflict of Interest

The authors declare that there are no conflicts of interest with respect to the authorship of this paper.

Authors Contribution

Conceptualization, BH and WR.; methodology, WR and MDKP.; software, MDKP.; formal analysis, WR and MDKP.; investigation, BH and WR.; data curation, IS, TD and VGSP.; writing—original draft preparation, BH., WR and MDKP; writing—review and editing, BH, WR and MDKP.; visualization, IS, TD and VGSP.; supervision, BH and WR. KI writing-reviewing and editing. All authors have read and agreed to the published version of the manuscript

References

Adams, R. J., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21*(1), 1-23. https://doi.org/10.1177/0146621697211001

Akhidime, A. E. (2017). The importance and development of research problem: A didactic discuss. *International Journal of Economics, Commerce and Management, 5*(8), 631-640.

American Educational Research Association (AERA), American Psychological Association (APA), National Council for Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Anthonysamy, L. (2021). The use of metacognitive strategies for undisrupted online learning: Preparing university students in the age of pandemic. *Education and Information Technologies.* Advanced online publication. https://doi.org/10.1007/s10639-021-10518-y

Arianto, M. A., & Wulyani, A. N. (2022). Self-regulated writing (SRW) strategies during the completion of dissertation: A case study. *Cakrawala Pendidikan, 41*(1), 154-169. https://doi.org/10.21831/cp.v41i1.40687

Baker, B. F., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). Marcel Dekker.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443–459. https://doi.org/10.1007/BF02293801

Bock, R.D., & Gibbons, R. D. (2021). *Item response theory*. Wiley.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a metropolis-hastings robbins-monro algorithm. *Psychometrika,* 75(1), 33-57. https://doi.org/10.1007/s11336-009-9136-x

Cer, E. (2019). The instruction of writing strategies: The effect of the metacognitive strategy on the writing skills of pupils in secondary education. *SAGE Open.* https://doi.org/10.1177/2158244019842681

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1-29. https://doi.org/10.18637/jss.v048.i06

Chou, Y.-T., & Wang, W.-C. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement*, *70*(5), 717–731. https://doi.org/10.1177/0013164410379322

Craig, K., Hale, D., Grainger, C., & Stewart, M. E. (2020). Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning,* 15, 155-213. https://doi.org/10.1007/s11409-020-09222-y

Daniel, F., & Taneo, P. N. L. (2019). Analisis kesulitan mahasiswa dalam penyusunan proposal penelitian pendidikan matematika. *Jurnal Pendidikan Matematika Indonesia, 4*(2), 79-83.

De Ayala, R. J. (1994). The influence of multidimensionality on the graded response model. *Applied Psychological Measurement, 18*(2), 155-170. https://doi.org/10.1177/014662169401800205

De Ayala, R. J. (2022). *The theory and practice of item response theory*. Guilford Press.

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities*, *34*(5), 435–449. https://doi.org/10.1177/002221940103400505

Dhir, S. K., & Gupta, P. (2021). Formulation of research question and composing study outcomes and objectives. *Indian Pediatrics,* 58, 584-588. https://doi.org/10.1007/s13312-021-2246-y

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modelling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, *16*(1), 5–18. https://doi.org/10.1007/s11136-007-9198-0

Faculty of Psychology University of Indonesia. (2020). *Buku pendoman akademik program studi doktor psikologi*. Faculty of Psychology University of Indonesia.

Farrugia, P., Petrisor, B. A., Farrokhyar, F., & Bhandari, M. (2010). Research questions, hypotheses and objectives. *Canadian Journal of Surgery, 53*(4), 278-281.

Faustino, B., Vasco, A. B., Oliveira, J., Lopes, P., & Fonseca, I. (2021). Metacognitive self-assessment scale: Psychometric properties and clinical implications. *Applied Neuropsychology: Adult*, *28*(5), 596–606. https://doi.org/10.1080/23279095.2019.1671843

Firdayani., Fauzan., Nahartini, D., Yuliani, A., & Latip, A. (2020). Analysis of goals, metacognition and academic achievement students' of Muhammadiyah Malang University. In T. Hariguna & S.

C. Chen (Eds.), *Proceedings of the 1st International Conference on Recent Innovations* (ICRI 2018, pp. 1248-1256). Science and Technology Publications, Lda. http://doi.org/10.5220/0009925712481256

Flavell, J. (1979). Metacognition and cognitive monitoring. A new area of cognitive development inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906.

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert, & R. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 21-29). Lawrence Erlbaum Associates.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement, 21*(4), 347-360. https://doi.org/10.1111/j.1745-3984.1984.tb01039.x

Hameed, H. A., & Cheruvalath, R. (2021). Metacognitive skills inventory (msi): Development and validation. *International Journal of Testing*. Advance online publication. https://doi.org/10.1080/15305058.2021.1986051

Inouye, K., & McAlpine, L. (2022). Writing across contexts: Relationships between doctoral writing and workplace writing beyond the academy. *Innovations in Education and Teaching International*. Advanced online publication. https://doi.org/10.1080/14703297.2022.2075431

Jans-Beken, L., & Wong, P. T. P. (2021). Development and preliminary validation of the existential gratitude scale (EGS). *Counselling Psychology Quarterly, 34*(1), 72–86. https://doi.org/10.1080/09515070.2019.1656054

Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. Structural Equation Modeling, *15*(1), 136–153. https://doi.org/10.1080/10705510701758406

Kang, T., & Chen, T. (2008). Performance of the generalized S - χ^2 item fit index for polytomous IRT models. *Journal of Education Measurement*, *45*(4), 391–406. https://doi.org/10.1111/j.1745-3984.2008.00071.x

Kircaburun, K., Stravropoulos, V., Harris, A., Calado, F., Emirtekin, E., & Griffiths, M. D. (2021). Development and validation of the mukbang addiction scale. *International Journal of Mental Health and Addiction, 19*, 1031-1044. https://doi.org/10.1007/s11469-019-00210-1

Kocimaheni, A. A., Aminin, Z., & Kartika, A. D. (2020). Identifikasi kesulitan mahasiswa fakultas bahasa dan seni universitas negeri surabaya dalam penyelesaian studinya. *Paramasastra, 7*(2), 139-146.

Kuiken, F., & Vedder, I. (2021). The interplay between academic writing abilities of Dutch undergraduate students, a remedial writing programme, and academic achievement. *International Journal of Bilingual Education and Bilingualism*, *24*(10), 1474-1485. https://doi.org/10.1080/13670050.2020.1726280

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936-949. https://doi.org/10.3758/s13428-015-0619-7

Marhaban, S., Mukminatien, N., Widiati, U., Sulistyo, T., Suhastyanang, W. D., Puspitasari, Y., & Muslem, A. (2021). Strategies employed by EFL doctoral candidates in dissertation writing completion. *Studies in English Language and Education, 8*(2), 546-560. https://doi.org/10.24815/siele.v8i2.17694

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275–340). Lawrence Erlbaum Associates Publishers.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 49-174. https://doi.org/10.1007/BF02296272

Maydeu-Olivares, A. (2013). Why should we assess the goodness-of-fit of IRT models? *Measurement: Interdisciplinary Research and Perspectives*, *11*(3), 127-137. https://doi.org/10.1080/15366367.2013.841511

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*(4), 713-732. https://doi.org/10.1007/s11336-005-1295-9

Mya, K. S., Zaw, K. K., & Mya, K. M. (2021). Developing and validating a questionnaire to assess an individual's perceived risk of four major non-communicable diseases in Myanmar. *PLoS ONE*, *16*(4), e0234281. https://doi.org/10.1371/journal.pone.023428

Mbato, C. L., & Cendra, A. (2019). EFL undergraduate students' self-regulation in thesis writing: Help-seeking and motivation-regulation. *Journal of English Language and Education, 5*(1), 66-82. http://dx.doi.org/10.26486/jele.v5i1.949

Newman, I., & Covrig, D. M. (2013). Building consistency between title, problem statement, purpose, & research questions to improve the quality of research plans and reports. *New Horizons in Adult Education and Human Resource Development, 25*(1), 70-79. https://doi.org/10.1002/nha.20009

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: A meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacognition and Learning*, *13*(2), 179–212. https://doi.org/10.1007/s11409-018-9183-8

Papageorgiou, C., & Wells, A. (2003). An empirical test of a clinical metacognitive model of rumination and depression. *Cognitive Therapy and Research*, *27*(3), 261–273. https://doi.org/10.1023/A:1023962332399

Pedone, R., Semerari, A., Ricciardi, I., Procacci, M., Nicolo, G., & Carcione, A. (2017). Development of a self-report measure of metacognition: The metacognition self-assessment scale (MSAS): Instrument description and factor structure. *Clinical Neuropsychiatry, 14*(3), 185-194.

Petscher, Y., Mitchell, A. M., & Foorman, B. R. (2015). Improving the reliability of student scores from speeded assessments: an illustration of conditional item response theory using a computer-administered measure of vocabulary. *Reading and Writing*, *28*(1), 31-56. https://doi.org/10.1007/s11145-014-9518-z

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychological Review, 16*(4), 385-407. https://doi.org/10.1007/s10648-004-0006-x

Quvanch, Z., & Na, K. S. (2022). Evaluating Afghanistan university students' writing anxiety in English class: An empirical research. *Cogent Education, 9*(1), 2040697. https://doi.org/10.1080/2331186X.2022.2040697

Reise, S. P. (1999). Personality measurement issues viewed through the eyes of IRT. In S. E. & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 219–241). Erlbaum.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*(3), 287-297. https://doi.org/10.1037/1040-3590.12.3.287

Rifenbark, G., Lombardi, A., & Freeman, J. (2021). A confirmatory item factor analysis of a school climate measure for adolescents with and without disabilities. *Educational Assessment, 26*(1), 52-68. https://doi.org/10.1080/10627197.2020.1841625

Roth, A., Ogrin, S., & Schmitz, B. (2016). Assessing self-regulated learning in higher education: A systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability*, *28*(3), 225–250. https://doi.org/10.1007/s11092-015-9229-2

Saeed, M. A., Al-Ahdal, A. A. M. H., & Al Qunayeer, H. S. (2021). Integrating research proposal writing into a postgraduate research method course: what does it tell us? *International Journal of Research & Method in Education, 44*(3), 303-318. https://doi.org/10.1080/1743727X.2020.1777963

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *34*(4, Pt. 2), 100.

Song, J. H. H., Loyal, S., & Lond, B. (2021). Metacognitive awareness scale, domain specific (MCAS-DS): Assessing metacognitive awareness during Raven's Progressive Matrices. *Frontiers in Psychology,* 11:607577. https://doi.org/10.3389/fpsyg.2020.607577

Stover, A. M., McLeod, L. D., Langer, M. M., Chen, W-H., & Reeve, B. B. (2019). State of the psychometric methods: patient-reported outcome measure development and refinement using item response theory. *Journal of Patient-Reported Outcomes*, 3:50. https://doi.org/10.1186/s41687-019-0130-5

Su, S., Wang, C., & Weiss, D. J. (2021). Performance of the S - χ^2 statistic for the multidimensional graded response model. *Educational and Psychological Measurement, 81*(3), 491-522. https://doi.org/10.1177/0013164420958060

Suseno, M., Hayat, B., Putra, M. D. K., Bien, J. K., Rachmawati, R., & Hartanto, H. (2022). A differential item functioning (DIF) analysis of the mobile phone problem use scale in Indonesian schools with and without smartphone banned policy. *Cogent Psychology, 9*(1), 2137306. https://doi.org/10.1080/23311908.2022.2137306

Teng, F. (2020). The role of metacognitive knowledge and regulation in mediating university EFL learners' writing performance. *Innovation in Language Learning and Teaching, 14*(5), 436-450. https://doi.org/10.1080/17501229.2019.1615493

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: the test is not reliable. *Educational and Psychological Measurement*, *60*(2), 174-195. https://doi.org/10.1177/0013164400602002

Uesaka, Y., Suzuki, M., & Ichikawa, S. (2022). Analyzing students' learning strategies using item response theory: Toward assessment and instruction for self-regulated learning. *Frontiers in Education,* 7: 921884. https://doi.org/10.3389/feduc.2022.921844

van de Schoot, R., Yerkes, M. A., Mouw, J. M., & Sonneveld, H. (2013). What took them so long? Explaining phd delays among doctoral candidates. *PloS ONE, 8*(7), e68839. https://doi.org/10.1371/journal.pone.0068839

van Rooij, E., Fokkens-Bruinsma, M., & Jansen, E. (2021). Factors that influence phd candidates' success: the importance of PhD project characteristics. *Studies in Continuing Education, 43*(1), 48-67. https://doi.org/10.1080/0158037X.2019.1652158

Wolters, C. A., & Hussain, M. (2015). Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition and Learning*, *10*(3), 293–311. https://doi.org/10.1007/s11409-014-9128-9

Xia, W., & Luxin, Y. (2012). Problems and strategies in learning to write a thesis proposal: A study of six M.A. students in a TEFL program. *Chinese Journal of Applied Linguistics (Quarterly), 35*(3), 324-341. https://doi.org/10.1515/cjal-2012-0024

Yau, D. T. W., Wong, M. C. M., Lam, K. F., & McGrath, C. (2015). Evaluation of psychometric properties and differential item functioning of 8-item Child Perceptions Questionnaires using item response theory. *BMC Public Health,* 15:792. https://doi.org/10.1186/s12889-015-2133-3