Jurnal Elemen, 10(3), 614-629, September 2024
https:/doi.org/10.29408/jel.v10i3.26618

e- ISSN: 2442 - 4226

Check for updates

Utilizing item response theory for the analysis of self-regulated learning scale in mathematics

Sintha Sih Dewanti *, Jasmine Nurul Izzah, Shinta Puspa Kiranasari, Kholifatul Fatoni Sholihin

Department of Mathematics Education, UIN Sunan Kalijaga Yogyakarta, Yogyakarta, Indonesia

* Correspondence: sintha.dewanti@uin-suka.ac.id

Abstract

In the learning process, students must be able to regulate themselves. This is an effort to improve the quality of student learning, especially their learning achievement. The purpose of this study is to empirically analyze the characteristics of self-regulation skills in learning mathematics. The analysis was carried out by utilizing Item Response Theory (IRT) with a Partial Credit Model (PCM). This study is a descriptive quantitative study with the subjects of the study being 123 students of grade 10 of Senior High Schools in Yogyakarta. Data collection was carried out using a questionnaire that had been developed and its item characteristics were analyzed using IRT. The independent learning scale questionnaire in mathematics consists of 16 statement items and measures 5 aspects. Based on all the questions that are feasible to be analyzed using the PCM model, it is known that there are 5 aspects that fit, namely (1) self-confidence, (2) discipline in learning, (3) active in learning, (4) responsibility, and (5) motivation in learning. This instrument can be used to measure students' self-regulation skills in learning mathematics based on their ability level.

Keywords: item response theory; mathematics; partial credit model; self-regulated learning

How to cite: Dewanti, S. S., Izzah, J. N., Kiranasari, S. P, & Sholihin, K. F. (2024). Utilizing item response theory for the analysis of self-regulated learning scale in mathematics. *Jurnal Elemen*, *10*(3), 614-629. https://doi.org/10.29408/jel.v10i3.26618

Received: 19 June 2024 | Revised: 25 August 2024
Accepted: 24 September 2024 | Published: 1 October 2024

Introduction

Self-management skills are crucial skills that every individual has and are very important in living everyday life. According to Hidayat et al. (2020), this ability includes initiative, resilience in the face of obstacles, self-confidence, and the ability to act independently without the help of others. This ability is also often associated with self-control and regulation of behavior to achieve personal and professional goals (Zimmerman & Kitsantas, 2014; Panadero, 2017). More profoundly, self-regulation skills include focusing attention and self-control in dealing with various situations so that individuals do not experience emotional dependence on others. Zimmerman and Moylan (2009) emphasized that this ability is essential in lifelong learning, allowing individuals to learn effectively without direct supervision.

Recent research also shows that individuals with good self-regulation skills have better mental health and can cope with stress more effectively (Duckworth et al., 2016; Hoyle & Davisson, 2021). In addition, self-regulation is also associated with better academic performance, as individuals who can manage their time and responsibilities independently tend to be more successful in achieving academic goals (Perry et al., 2018). This ability allows one to face life's challenges more independently and without relying on others, significantly contributing to personal and professional success (Conti et al., 2023).

Self-regulated learning (SRL) is closely related to the ability to learn independently, where individuals do not rely on the help of others in their learning process. This ability allows students to take full responsibility for their learning, show strong willpower, and have high discipline, ultimately contributing to optimal learning outcomes (Asmar, 2018). Recent research also supports that students with good SRL abilities are more likely to achieve higher academic achievement because they can set goals, monitor progress, and manage their time and resources effectively (Panadero et al., 2020a; Perry et al., 2018). According to Zimmerman (1990), SRL differs from other learning approaches because it emphasizes how students actively choose, organize, and create meaningful learning environments and plan and regulate their behavior to achieve learning goals.

SRL also involves cognitive and metacognitive strategies, which help students plan, monitor, and evaluate their learning (Dinsmore & Alexander, 2016; Schunk & Greene, 2018). This strategy includes adapting to various learning challenges, both emotionally and cognitively, so that students can overcome obstacles and improve the quality of learning (de Bruijn-Smolders et al., 2016). In the context of education, the development of SRL in students is significant because it has been proven effective in improving learning outcomes, especially in increasingly autonomous and technology-based learning environments (Kitsantas et al., 2017; Broadbent & Poon, 2018). Therefore, the application of SRL must be the main focus of the learning process because it is not just a skill but also the foundation for long-term success in students' education and careers (Wolters & Brady, 2020). By developing SRL, students improve their learning achievement and prepare themselves to become lifelong learners.

SRL has a very close relationship with the ability to learn independently without relying on others. Learning that emphasizes independence allows students to take full responsibility for their learning process and develop a strong will and high discipline. It ultimately contributes to achieving optimal learning outcomes (Asmar, 2018). According to Zimmerman (1990), SRL differs from other learning methods because it focuses on how students actively choose, organize, and create a meaningful learning environment. Students also plan and regulate their behavior to achieve predetermined goals (Zimmerman, 1990). In the learning process, SRL is essential because it can improve students' quality of learning and overall academic achievement (Panadero et al., 2020b; Greene et al., 2018).

SRL also drives students to have intrinsic motivation and high curiosity about the ever-evolving science (Schunk & Greene, 2018). Thus, students are encouraged to explore new things in learning and develop their ability to learn independently without relying on external assistance (Broadbent & Fuller-Tyszkiewicz, 2018; Vahuka et al., 2023). Recent studies have shown that students with good SRL skills will be better able to acquire new knowledge widely and evaluate their abilities, ultimately improving their learning performance (Bembenutty et al., 2020; Panadero et al., 2019). In addition, with the knowledge obtained through SRL, students can independently achieve their learning goals and consistently improve their learning outcomes (Hodges & McLeod, 2020). This process improves academic achievement and helps students develop lifelong learning skills, which are very important in facing future educational and career challenges (Dinsmore & Alexander, 2020).

Students need to be aware of learning independently and be able to determine the steps that must be taken in their learning process. Indicators are needed to measure students' learning independence to support this. One way to evaluate learning independence instruments is through item analysis, which can be done with two main approaches: Classical Test Theory (CTT) and Item Response Theory (IRT). CTT is a traditional approach in measurement theory that explains the relationship between observed scores on tests and actual scores that cannot be measured directly (Wang & Osterlind, 2013). In measuring learning independence, CTT helps determine the extent to which the instrument can measure the desired domain. The advantages of CTT are that this method is easy to understand, has a simple analysis procedure, and only requires a relatively small sample size (Taherdoost, 2019). However, the weakness of CTT lies in its nature, which depends on the group of test takers (group dependent) and the characteristics of the test items (item dependent), as well as the estimation of the standard error of measurement (SEM) which applies equally to all test takers, making it less accurate in some cases (Dewanti et al., 2021).

As an alternative, IRT was developed to overcome the limitations of CTT with a more flexible approach. IRT builds a model that connects the characteristics of test items with the abilities of participants in more detail so that it can provide more accurate information about the quality of the questions and abilities (Raju et al., 2020). IRT in item analysis offers advantages over CTT, especially in estimating item and participant ability parameters more precisely and consistently (Santoso, 2018). By using IRT, researchers can obtain a more comprehensive picture of the characteristics of SRL in students so that the measurement instruments developed are more valid and reliable (Sarea & Ruslan, 2019).

Methods

This study was a descriptive quantitative study that empirically analyzed the characteristics of students' SRL based on the IRT approach. The population of this study consisted of 250 grade X students at one of the state high schools in Yogyakarta. From this population, the research sample was 123 students using a particular sampling technique. The research instrument used was a non-test questionnaire to measure self-regulated learning. The questionnaire consists of 20 statement items with five answer options: Always (A), Often (O), Sometimes (S), Ever (E), and Never (N). Data was collected by distributing questionnaires to students through the Google Forms platform to facilitate distributing and filling in online (Ghergulescu & Muntean, 2021). The IRT approach in data analysis allows researchers to identify the characteristics of test items and respondents' abilities more accurately than traditional measurement methods (Wilson, 2018).

According to Hendriana et al. (2017), indicators of SRL include (1) not depending on others, (2) self-confidence, (3) disciplined behavior, (4) having initiative, (5) self-control, and (6) having a sense of responsibility. Meanwhile, indicators of SRL, according to Zimmerman and Schunk (2001), include: (1) metacognition (planning their learning activities, determining goals in learning, monitoring themselves when learning, and evaluating themselves in learning), (2) motivation (efforts to do better learning, having self-efficacy in their learning activities), and (3) behavior (efforts to determine and create an environment that can optimize their learning activities). Mudjiman (2011) describes indicators of self-regulated learning, namely (1) self-confidence, (2) discipline in learning, (3) activity in learning, (4) responsibility, and (5) motivation in learning. The questionnaire used in this study was prepared by referring to the indicators of SRL according to Hendriana et al. (2017), which consists of (1) not dependent on others, (2) self-confidence, (3) disciplined behavior, (4) having their initiative, (5) self-control, and (6) having a sense of responsibility.

Each aspect is measured with three to four items. The indicator of not depending on others assesses the aspects of learning on their own, being able to solve problems without the help of others, being able to make their own decisions, and determining learning goals/strategies. The indicator of self-confidence assesses believing in their abilities, participating in discussion activities, and not cheating on tests. The indicator of disciplined behavior assesses obeying the rules during learning, collecting assignments on time, and being afraid when breaking the rules. Self-initiative indicators assess looking for other learning resources, understanding learning needs, and looking for alternatives in solving problems. Indicators of self-control assess aspects of knowing strengths and weaknesses, knowing material that is considered difficult, and enthusiasm for learning. Indicators of a sense of responsibility assess commitment to learning, earnestness, and focus.

The questionnaire scale was validated by 6 validators, including mathematics education lecturers, mathematics teachers, and prospective mathematics teachers. Validators evaluated the questionnaire on three aspects: presentation, material, and language. All items were deemed valid, as evidenced by Aiken's V coefficient value ≥ 0.75, with six validators and 5 answer choices (Aiken, 1985).In terms of measurement error, the instrument's reliability is demonstrated by Cronbach's Alpha coefficient of 0.445, indicating sufficient reliability. This high level of reliability ensures the security of the research findings and the trustworthiness of the questionnaire.

In this study, the data from the test results of the SRL questionnaire instrument were analyzed using IRT with the help of the SPSS version 23, LISREL 8.80, and RStudio programs. SPSS is used to test factor analysis, namely Exploratory Factor Analysis (EFA). Factor analysis is divided into EFA and Confirmatory Factor Analysis (CFA). CFA was analyzed using the LISREL program in this study. At the same time, the R program was used for data analysis of the IRT model with the PCM approach.

Results

Analysis with IRT is carried out in two stages, namely the assumption test stage and the test analysis stage. In analyzing data using IRT, the first thing to do is to test the dimensions of empirical data. The testing process is done with EFA using the principal component method. The initial assumptions that must be met in EFA are the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett tests. The KMO and Bartlett test output is presented in Table 1.

Table 1. KMO and Bartlett test output

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.780
Bartlett's Test of Sphericity	Approx. Chi-Square	716.123
	df	190
	Sig.	.000

Based on the results of the KMO and Bartlett tests, the data were analyzed to determine the number of factors formed. The number of factors can be determined by selecting factors that have Eigenvalues greater than 1. The extraction process is carried out to obtain items that measure the same dimensions so that several factors are produced. The scree plot of the number of factors formed can be seen in Figure 1.

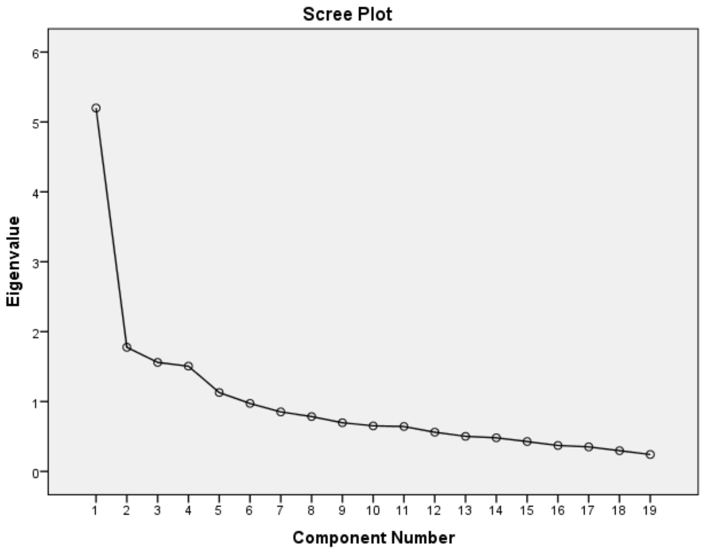

Figure 1. Scree Plot

After new factors are formed through factor analysis, the SPSS output will display a rotated component matrix, which shows how strongly the items in the instrument correlate with the resulting factors. Typically, an item is considered to have a significant contribution to a factor if its factor loading value is greater than 0.50 (Hair et al., 2019). If the items have a factor loading value above 0.50 and are consistently grouped on one factor, then it can be concluded that the items theoretically represent the same dimension or factor (Field, 2018). This rotation process helps clarify the pattern of relationships between items, facilitates the interpretation of factors, and ensures that the formed factors more accurately reflect the underlying latent structure of the data (Beavers et al., 2017). Grouping items based on factors can be seen in Table 2.

Table 2. Rotated component matrix

	Component				
	1	2	3	4	5
i18	.801				
i19	.789				
i20	.636				
i10	.604				
i5		.749			
i2		.720			
i3		.603			
i11			.809		
i1			.667		
i17			.571		
i15			.566		
i6				.709	
i14				.700	
i12				.595	
i9					.842
i16					.505

	Component				
	1	2	3	4	5
Extraction Method: Principal Component Analysis Rotation Method: Varimax with Kaise Normalization					
a. Rotation converged in 8 iterations					

The Rotated Component Matrix output in factor analysis aims to clarify the relationship between variables and the resulting factors. Rotation is done to facilitate interpretation, with the most used rotation method being Varimax, which is an orthogonal rotation method (Hair et al., 2019). In this matrix, the factor loading value indicates the strength of the relationship between each item and the factors formed. Generally, items with a factor loading value greater than 0.50 are considered to have a strong correlation with the factor, and this value is often used as a limit to determine whether an item can be retained in the factor (Brown, 2015).

The rotation process reduces the complexity of the results and helps identify items exclusively related to one factor, thereby improving the readability and interpretation of the factor structure (Howard, 2016). It is essential because items may correlate with multiple factors without rotation, making interpretation difficult. Therefore, rotation provides a solution to minimize cross-loading and ensure each item is related to one dominant factor (Williams et al., 2020). Using a Rotated Component Matrix is essential in psychometric research or educational surveys because it helps ensure that the items in the measurement instrument measure the intended construct (Costello & Osborne, 2019). Thus, rotation improves the interpretability of the results, and the validity of the factors formed from the data. The construct validity of SRL was proved by confirmatory factor analysis. The CFA output is presented in Figure 2.

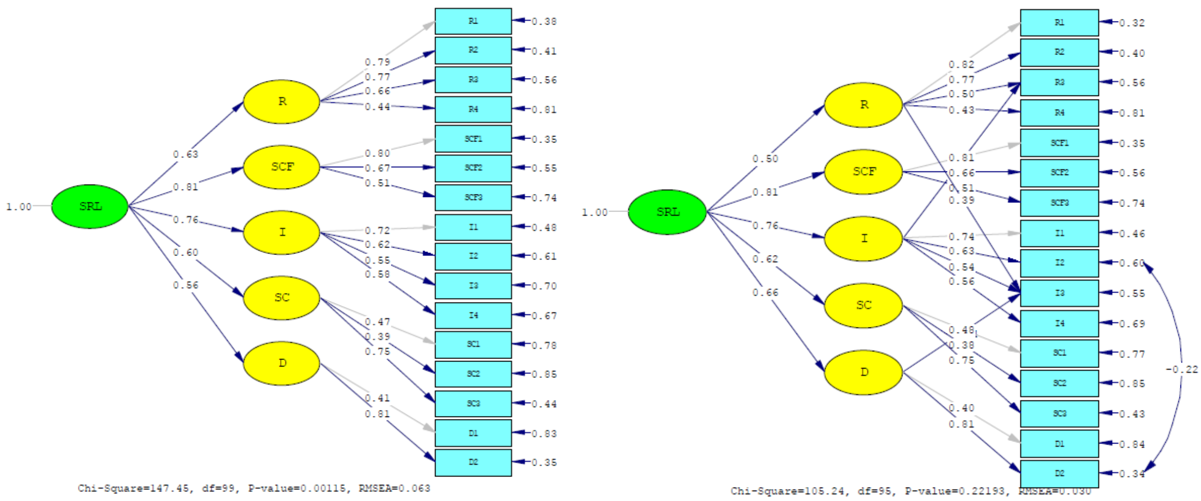

Figure 2. (a) CFA output before modification; (b) CFA output after modification

Based on the results of the CFA before modification, the p-value obtained has not reached a value > 0.05, which indicates that the model does not meet the goodness-of-fit criteria. This indicates the need for further review of the modification index to identify parts of the model that need to be repaired or changed (Byrne, 2016). The modification index provides suggestions for improvements that can improve the suitability of the model, for example by increasing the correlation between errors or removing irrelevant items (Schumacker & Lomax, 2018).

After the model modification was carried out, the CFA results showed that the p-value increased to 0.22193, which exceeded the threshold of 0.05. This indicates that the modified model fits the estimated population covariance and the sample data covariance better, so the model can be considered statistically fit (Kline, 2020). A p-value greater than 0.05 means no significant difference exists between the covariance matrix estimated by the model and that found in the sample data, indicating that the model is suitable for further hypothesis testing (Brown, 2015). With appropriate modifications, the model can meet various goodness-of-fit criteria, such as CFI, TLI, RMSEA, and SRMR, which must also be considered to ensure the validity and reliability of CFA results (Hair et al., 2019).

Before proceeding to the next test, conducting a model fit test is crucial. This step is essential to determine whether the data aligns with the criteria of the existing research model (Nurbaiti, 2021). It ensures that the theoretical model is not just a concept but, in fact, the empirical data. The criteria for testing the Goodness of Fit Model are presented in Table 3.

Table 3. Goodness of fit model

No.	Statistics	Result	‘”Fit” Criteria	Description
1	P-value	0.22193	> 0.05	Fit
1	RMSEA	0.030	< 0.08	Fit
2	GFI	0.90	≥ 0.90	Fit
3	CFI	0.97	≥ 0.90	Fit
4	AGFI	0.86	0.80≤ AGFI < 0.9	Marginal fit

The results of the PCM analysis using the R program provide output in the form of an Item Characteristic Curve (ICC) for each question item. The ICC plot illustrates the relationship between the respondent's ability level and the probability of answering correctly on a particular question item (De Ayala, 2013). For good questions, the ICC curve shows a clear and sequential pattern, where the intersection between the n category function and the n + 1 category function occurs systematically and does not overlap. The further to the right of the plot, the level of difficulty of the question (b) increases, meaning that respondents with higher abilities are more likely to answer the question correctly (Embretson & Reise, 2013).

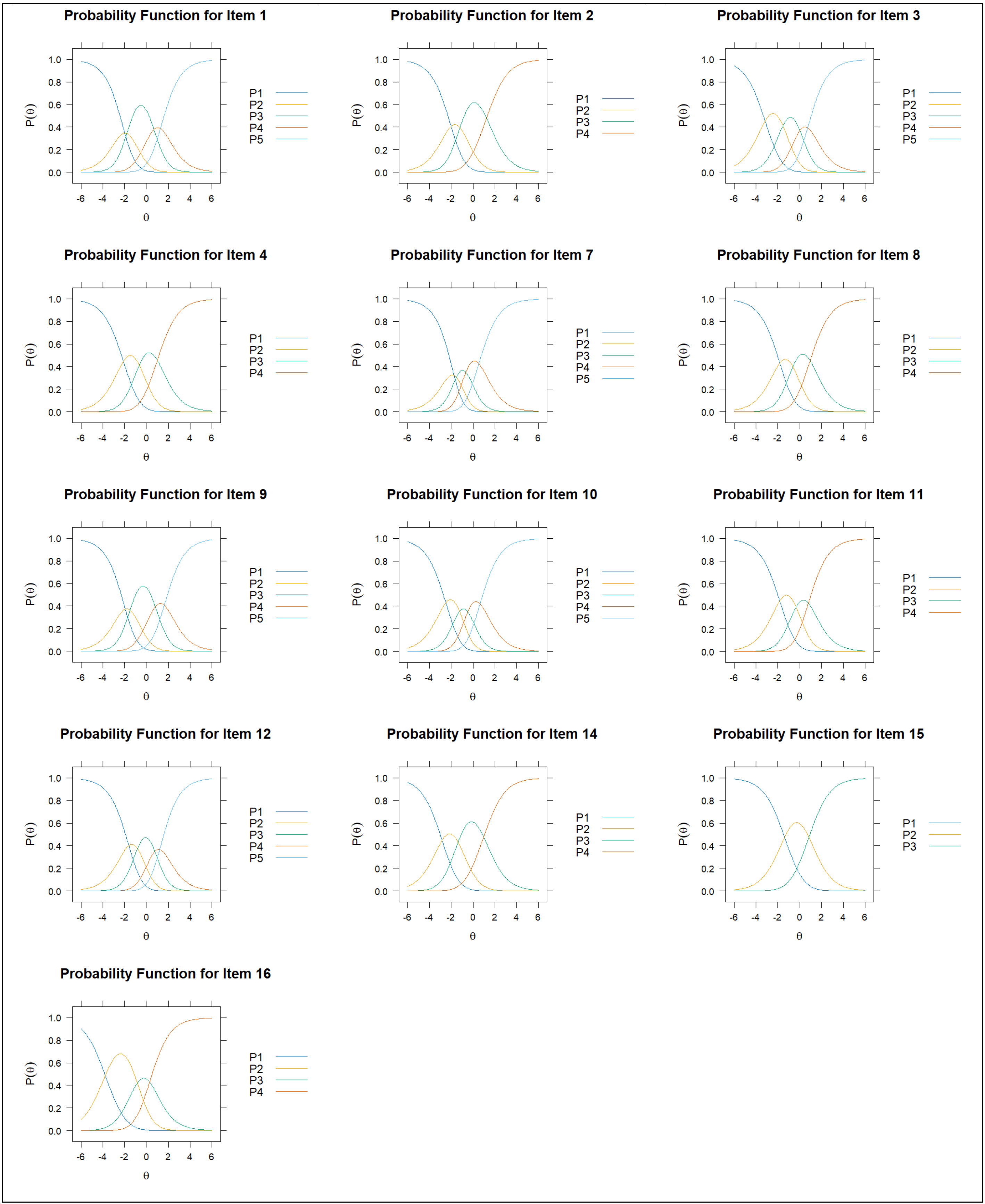

Figure 3. Ideal ICC plot

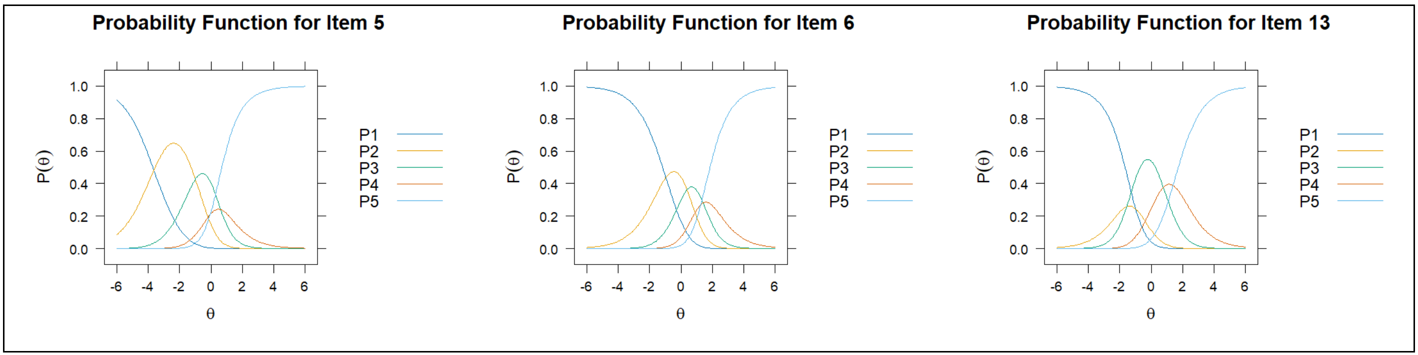

Figure 4. Not ideal ICC plot

An ideal ICC plot is characterized by neat and sequential lines with consistent distances between categories. In this plot, each response category has a clear area according to the ability level being measured, so the item can accurately distinguish respondents with different levels of ability (van der Linden, 2016). Conversely, a non-ideal ICC plot shows overlap between category functions, where the response lines are not sequential, or some categories do not have clear areas. It suggests that the item is ineffective in differentiating respondents based on their ability levels and requires modification or deletion (Hambleton et al., 2019). Non-ideal ICC plots can be caused by various factors, including ambiguity in the item or difficulty levels that do not match the average ability of respondents (Bond & Fox, 2015).

Discussion

This study aims to empirically analyze the characteristics of SRL of high school students using the IRT approach. In research, it is essential to develop an effective instrument that is easy to use and provides accurate and reliable results for analysis purposes (Dewanti et al., 2020). The instrument that has been designed must go through validity testing to ensure that the measuring instrument can measure what should be measured. Validity is the primary measure that determines the instrument's accuracy level in measuring the variables studied (Hair et al., 2019). Validity testing is carried out by considering three main aspects: criterion validity, content validity, and construct validity. Construct validity examines the extent to which an instrument can measure the expected theoretical concept (Taherdoost, 2016). In this study, construct validity was assessed through a series of methods, including assessment by validators, Aiken coefficients to measure agreement between experts, and EFA and CFA analyses to test the instrument's factor structure (Kline, 2020).

The PCM is a polytomous model within IRT and is an extension of the Rasch model. PCM assumes that each item has the same discrimination power but differs in difficulty levels (Masters, 1982). The primary focus in PCM is on estimating item difficulty parameters, which provide insights into the thresholds between different response categories experienced by participants (De Ayala, 2013). An essential assumption of PCM is that while the discrimination between items remains constant, the ordering of difficulty levels across response categories does not have to be strictly linear. It means a given category may be more difficult or accessible than the preceding or following one (Bond & Fox, 2015). Therefore, higher scores on the PCM instrument indicate a higher level of ability in the participant than lower scores (van der Linden & Hambleton, 2017).

The results of the Kaiser-Meyer-Olkin (KMO) test in Table 1 show a KMO value of 0.780, which is greater than 0.5, indicating that the sample size is sufficient for factor analysis (Field, 2018). In addition, the results of the Bartlett's Test of Sphericity provide a significance value of 0.000 (<0.05), indicating that the null hypothesis (the correlation matrix is an identity matrix) is rejected. This means that the correlation matrix between variables has a close enough relationship for factor analysis (Hair et al., 2019). Based on the Anti-image Correlation value, it is known that the Measures of Sampling Adequacy (MSA) for each item is above 0.50, except for item 13. Therefore, all items (except item 13) are considered worthy of further factor analysis (Tabachnick & Fidell, 2019). Item 13 was excluded from the analysis because the MSA value was below 0.50, indicating that the item was not suitable for inclusion in the analysis (Fabrigar & Wegener, 2018).

Based on the KMO and Bartlett's tests, the initial analysis did not meet all the necessary conditions for factor analysis due to the MSA (Measure of Sampling Adequacy) value for item 13 being 0.440 (MSA < 0.50). Therefore, re-analysis was required after eliminating item 13. Following this adjustment, the KMO value improved to 0.792, and Bartlett's Test of Sphericity yielded a significant value ($p < 0.05$), indicating that the data were suitable for factor analysis (Hair et al., 2019). Furthermore, the MSA values for all remaining items exceeded 0.50, confirming the adequacy of the data for further analysis (Tabachnick & Fidell, 2019). The number of factors was determined by selecting those with Eigenvalues greater than 1. As shown in the scree plot (Figure 1), five factors emerged based on the Eigenvalue criteria, which is a common method for identifying the optimal number of factors in EFA (Field, 2018).

Table 2 presents the Rotated Component Matrix, which illustrates the grouping of items based on the five factors identified in the analysis. Items i18, i19, i20, and i10 have factor loadings greater than 0.50 and are grouped under a single factor (Component 1), indicating that these variables form Factor 1. Similarly, items i5, i2, and i3, with factor loadings above 0.50, cluster together under Component 2, thus forming Factor 2. Items i11, i1, i17, and i15 also have factor loadings greater than 0.50 and are grouped under Component 3, representing Factor 3. For Component 4, items i6, i14, and i12 have significant factor loadings (>0.50), leading to the creation of Factor 4. Lastly, items i9 and i16 are grouped under Component 5 with factor loadings exceeding 0.50, defining Factor 5 (Field, 2018; Tabachnick & Fidell, 2019). These five factors were subsequently labeled based on the underlying items: Factor 1 was named "Responsibility," Factor 2 as "Self-confidence," Factor 3 as "Initiative," Factor 4 as "Self-control," and Factor 5 as "Discipline" (Hair et al., 2019). The results from this EFA indicate that the questionnaire instrument is valid and reliable for measuring SRL (Henson & Roberts, 2016).

Item analysis was conducted using the PCM in RStudio. The output, including the Item Characteristic Curve (ICC) plots, is presented in Table 4. An ideal ICC plot is characterized by the intersection of curves between adjacent response categories, where the intersection shifts progressively to the right. This signifies that the thresholds between categories are ordered appropriately, reflecting increasing levels of difficulty (Bond & Fox, 2015). Nine items were found to have five thresholds, six items had four thresholds, and one item had three thresholds. Ideally, the thresholds should increase sequentially (e.g., threshold 1 < threshold 2 < threshold 3 < threshold 4 < threshold 5), as selecting "always" should indicate a higher propensity than selecting "often" (van der Linden & Hambleton, 2017). Thirteen items exhibited this increasing threshold pattern, indicating a well-functioning item structure. However, three items (Item 5, Item 6, and Item 13) displayed non-ideal threshold patterns, where a category threshold was not greater than the previous one (Andrich, 2018). Based on Tables 4 and 5, 13 items from the SRL instrument were deemed ideal, while 3 items were flagged for further review due to suboptimal threshold ordering (Masters & Wright, 2020).

Conclusion

A questionnaire instrument was developed to measure students' self-regulated learning. The validity test conducted included construct validity analyzed through the Aiken coefficient, EFA, CFA, and analysis using RStudio. The results of the Kaiser-Meyer-Olkin (KMO) test showed a value of 0.792, while Bartlett's Test of Sphericity produced a significant value of 0.000 (<0.05), which indicated the adequacy of the sample and the existence of a relationship between variables. All items showed a Measure of Sampling Adequacy (MSA) value above 0.50, so all data were worthy of further analysis. Based on the Eigenvalue value, five factors were formed, namely responsibility, self-confidence, initiative, self-control, and discipline. After being analyzed using the PCM, it was found that there were 9 items with 5 thresholds, 6 items with 4 thresholds, and 1 item with 3 thresholds. Overall, 13 items were considered ideal, while 3 items were considered not meeting the ideal criteria.

This study has several limitations, including the limited number of research subjects that can affect the generalizability of the results and the use of only one type of IRT model (PCM), which may need more to comprehensively evaluate all aspects of the test items. In addition, this instrument was only tested on one student population, so the results may not apply to student populations in different educational settings. The instrument that has been developed can be a valid and reliable tool for measuring SRL in students. However, further research is needed to expand the trial of this instrument on a more diverse population and add other analysis methods to ensure more substantial validity and reliability. In addition, further development of non-ideal test items can also improve the quality of this instrument.

Acknowledgment

We would like to thank the Faculty of Tarbiyah and Teacher Training of UIN Sunan Kalijaga Yogyakarta and the Senior High School where we conducted our research for providing the necessary facilities and opportunities to complete this research.

Conflicts of Interest

The authors declare no conflict of interest regarding the publication of this manuscript.

Funding Statement

This work received no specific grant from any public, commercial, or not-for-profit funding agency.

Author Contributions

Sintha Sih Dewanti: Research concept, article concept, writing and reviewing the article, validation and supervision, author correspondence; **Jasmine Nurul Izzah:** Writing–original draf, discussion, software, translation; **Shinta Puspa Kiranasari:** Data collection, analysis, methodology, translation; **Kholifatul Fatoni Sholihin:** Analysis, software, editing and visualization.

References

Aiken, L. R. (1985). Three coefficients for analyzing the reability and validity of ratings. *Educational and Psychological Measurement*, *45*(1), 131–142.

Asmar, E. (2018). Pengaruh kemandirian belajar dan minat belajar terhadap prestasi belajar ilmu pengetahuan alam [The influence of learning independence and learning interest on natural science learning achievement]. *Alfarisi: Jurnal Pendidikan MIPA*, *1*(1), 33–45.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2017). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, 18(1), 6.

Bembenutty, H., White, M. C., & DiBenedetto, M. K. (2020). *The roles of self-regulated learning and motivation in complex learning environments: Implications for college students*. In H. Bembenutty, T. J. Cleary, & A. Kitsantas (Eds.), Applications of Self-Regulated Learning across Diverse Disciplines (pp. 1-16). Information Age Publishing.

Bond, T. G., & Fox, C. M. (2015). *Applying the rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Broadbent, J., & Fuller-Tyszkiewicz, M. (2018). Self-regulated learning strategies and academic success in online higher education learning environments: A systematic review. *The Internet and Higher Education, 39*, 1-13.

Broadbent, J., & Poon, W. L. (2018). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 36*, 1-15.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.

Conti, R., Collins, M. A., & Picariello, M. L. (2023). Self-regulation in challenging tasks: Strategies and outcomes in modern settings. *Journal of Applied Psychology, 108*(3), 543–556.

Costello, A. B., & Osborne, J. W. (2019). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(1), 7.

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

de Bruijn-Smolders, M., Timmers, C. F., Gawke, J. C., Schoonman, W., & Born, M. P. (2016). *Effective self-regulatory processes: The role of metacognitive awareness and academic motivation in distance learning*. Learning and Individual Differences, 45, 223-230.

Dewanti, S. S., Ayriza, Y., & Setiawati, F. A. (2020). The application of item response theory for development of a students' attitude scale toward mathematics. *New Educational Review*, *60*, 108–123.

Dewanti, S. S., Hadi, S., Nu'man, M., & Ibrahim. (2021). The application of item response theory in analysis of characteristics of mathematical literacy test items. *Ilkogretim Online - Elementary Education*, *20*(1), 1226–1237.

Dinsmore, D. L., & Alexander, P. A. (2016). A critical discussion of deep and surface processing: What it means, how it is measured, the role of context, and model specification. *Educational Psychology Review, 28*(4), 1-29.

Dinsmore, D. L., & Alexander, P. A. (2020). *Self-regulation and learning: The interactive role of cognition, motivation, and emotions*. In J. Dunlosky & K. A. Rawson (Eds.), The Cambridge Handbook of Cognition and Education (pp. 223-246). Cambridge University Press.

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science, 11*(1), 35–55.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Psychology Press.

Fabrigar, L. R., & Wegener, D. T. (2018). *Exploratory factor analysis*. Oxford University Press.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.

Ghergulescu, I., & Muntean, C. H. (2021). Item Response Theory-based personalized self-regulated learning. *IEEE Transactions on Learning Technologies*, *14*(2), 134-147.

Greene, J. A., Hutchison, L. A., Costa, L.-J., & Crompton, H. (2018). Academic motivation and self-regulated learning in an educational technology course. *Computers & Education, 125*, 69-81.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (2019). *Fundamentals of item response theory*. SAGE Publications.

Hendriana, H., Rohaeti, E. E., & Sumarmo, U. (2017). *Hard skills dan soft skills matematika siswa*. Refika Aditama.

Henson, R. K., & Roberts, J. K. (2016). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement*, 76(3), 392–405.

Hidayat, D. R., Rohaya, A., Nadine, F., & Ramadhan, H. (2020). Kemandirian belajar peserta didik dalam pembelajaran daring pada masa pandemi COVID-19 [Learning independence of students in online learning during the COVID-19 pandemic]. *Perspektif Ilmu Pendidikan*, *34*(2), 147–154.

Hodges, C. B., & McLeod, S. (2020). Self-regulated learning and the importance of learner agency in online learning environments. *Distance Education, 41*(4), 477-488.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1), 51-62.

Hoyle, R. H., & Davisson, E. K. (2021). Self-regulation and well-being: An integrative perspective. *Annual Review of Psychology, 72*(1), 243–268.

Kitsantas, A., Winsler, A., & Huie, F. (2017). Self-regulation and motivation predictors of college student's academic performance and satisfaction. *Journal of Advanced Academics, 29*(3), 1-22.

Kline, R. B. (2020). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.

Masters, G. N., & Wright, B. D. (2020). The Partial Credit Model. In M. von Davier (Ed.), *Handbook of Item Response Theory* (Vol. 2, pp. 101-120). CRC Press.

Mudjiman. (2011). *Belajar mandiri [Independent learning]*. UNS Press.

Nurbaiti, B. (2021). *Tutorial pengolahan data dengan lisrel [Data processing tutorial with Lisrel]*. Pena Persada.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422.

Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Self-regulated learning in higher education: Strategies, challenges, and opportunities. *Educational Psychologist, 54*(4), 219-241.

Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2020a). Using technology to enhance self-regulated learning in higher education: The role of feedback and assessment. *Educational Technology Research and Development, 68*(2), 293-311.

Panadero, E., Jonsson, A., & Strijbos, J. W. (2020b). Scaffolding self-regulated learning through peer assessment: Opportunities and challenges. *Assessment & Evaluation in Higher Education, 45*(4), 545-560.

Perry, N. E., Rahim, A., & MacDonald, C. A. (2018). Self-regulated learning in the classroom: Realizing the potential for Canadian students. *Canadian Journal of Education, 4*1(3), 784–815.

Raju, D., Collins, M., & Cartwright, A. (2020). Item response theory and classical test theory: A comparative study in educational research. *Journal of Educational Measurement, 57*(1), 45-65.

Santoso, A. (2018). Karakteristik butir tes pengantar statistika sosial berdasarkan teori respon butir [Characteristics of introductory social statistics test items based on item response theory]. *Jurnal Pendidikan Matematika dan Sains*, *VI*(2), 1–11.

Sarea, M. S., & Ruslan, R. (2019). Karakteristik butir soal: Classical test theory vs item response theory? [Characteristics of test items: Classical test theory vs item response theory?]. *Didaktika: Jurnal Pendidikan*, *13*(1), 1–16.

Schumacker, R. E., & Lomax, R. G. (2018). *A beginner's guide to structural equation modeling* (4th ed.). Routledge.

Schunk, D. H., & Greene, J. A. (Eds.). (2018). *Handbook of self-regulation of learning and performance* (2nd ed.). Routledge.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics* (7th ed.). Pearson.

Taherdoost, H. (2016). Validity and reliability of the research instrument: How to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management*, 5(3), 28-36.

Taherdoost, H. (2019). Classical test theory vs. item response theory: A comparative analysis in educational assessments. *Educational Review, 31*(2), 12-28.

Vahuka, A. K., Sumartono, B. G., & Kuswanti, H. (2023). Analisis kemandirian belajar siswa pada mata pelajaran ekonomi di SMA Negeri 1 Teriak [Analysis of student learning independence in economics subjects at SMA Negeri 1 Teriak]. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, *12*(3), 855–862.

van der Linden, W. J. (2016). *Handbook of item response theory*. CRC Press.

van der Linden, W. J., & Hambleton, R. K. (2017). *Handbook of modern item response theory*. Springer.

Wang, Z., & Osterlind, S. J. (2013). Classical test theory. In *Handbook of Quantitative Methods for Educational Research* (pp. 31–44). Sense Publishers.

Williams, B., Onsman, A., & Brown, T. (2020). Exploratory factor analysis: A five-step guide for novices. *Journal of Emergency Primary Health Care*, 19(2), 42-49.

Wilson, M. (2018). Assessment with item response theory in educational research: A framework for applied practice. *Journal of Educational Measurement*, 55(3), 234-256.

Wolters, C. A., & Brady, A. C. (2020). College students' time management: A self-regulated learning perspective. *Educational Psychology Review, 32*(1), 131-154.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3–17.

Zimmerman, B. J., & Kitsantas, A. (2014). The role of self-regulated learning in collegiate achievement: A programmatic research and its implications. *Educational Psychologist, 49*(3), 210–226.

Zimmerman, B. J., & Moylan, A. R. (2009). *Self-regulation: Where metacognition and motivation intersect*. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 299–315). Routledge.

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement*. Taylor & Francis Groups.