

Expert Judgement of Self-Regulated Learning Questionnaire Quality: A Many-Facet Rasch Model Analysis

Rahmi Ramadhani^{1,2}, Edi Syahputra³, Elmanani Simamora⁴

{rahmiramadhani3@gmail.com^{1,2}, edisyahputra01.es@gmail.com³, elmanani_simamora@unimed.ac.id⁴}

Universitas Potensi Utama, Jl. K.L. Yos Sudarso KM. 6,5 No. 3-A Tanjung Mulia Medan¹,
Universitas Negeri Medan, Jl. Williem Iskandar Psr V Medan², Universitas Negeri Medan, Jl. Williem
Iskandar Psr V Medan³, Universitas Negeri Medan, Jl. Williem Iskandar Psr V Medan⁴

Abstract. Self-regulated learning (SRL) refers to a situation in which students proactively engage in their own learning process, and it is linked to the metacognitive, motivational, and behavioral attributes of students. SRL is an inherent trait among students that enables them to optimize their learning capacity, especially in the realm of mathematics. To assist students in exploring their SRL, SRL questionnaires are required. The objective of this research is to assess and scrutinize the quality of the SRL questionnaire developed in accordance with Zimmerman's stages. The quality of the SRL Questionnaire is evaluated and analyzed according to the following criteria: gender bias, language appropriateness, indicators that correspond to Zimmerman's stages, construction and systematics, relevance and systematics. The Many-Facet Rasch Model was used to assess the quality of the SRL questionnaire. FACETS software was used to collect and code 270 data points. The findings revealed that all 50 statements in the SRL questionnaire were valid and reliable enough to be employed in field research. Other findings revealed that the self-efficacy subscale in the SRL questionnaire was the best. The linguistic appropriateness criterion is the most challenging assessment criterion for experts to evaluate and judge.

Keywords: Expert Judgement, Many-Facet Rasch Model, Self-Regulated Learning, Questionnaire.

1 Introduction

Improving students' mathematical abilities requires the support of internal factors in each student, so these internal factors need to be continuously developed. One of the internal factors that support students' mathematical abilities is self-regulated learning. Self-regulated learning or SRL is a state in which students actively participate in their own learning and is closely related to students' metacognitive, motivational, and behavioral aspects. SRL has been shown to influence student performance, achievement, and learning outcomes [1], [2]. Students' self-conditions that can be described as SRL include behavior, motivation, and

cognition that can control goals related to students' academic problems [3]. SRL also makes a positive contribution to the effectiveness of the implementation of learning [4]–[6].

The importance of SRL suggests that every teacher should measure changes and improvements in students' SRL. This is done to ensure that students' internal conditions, such as behavior, motivation, and cognition, can support the optimization of students' mathematical abilities. Therefore, the creation of an instrument to measure SRL is critical. A questionnaire is the type of non-test instrument used to measure SRL. One cycle that is used in creating tools to measure students' capacity for SRL is the Zimmerman cycle. The Zimmerman cycle consists of three phases and six subscales, such as forethought phase (subscales: planning and self-efficacy), performance phase (subscales: self-monitoring and self-effort), and self-reflection phase (subscales: self-evaluation, self-satisfaction) [7].

An instrument cannot be used in the field unless it has undergone an expert evaluation called content validation. Content validation is an assessment that is measured based on a theoretical rationale and conducted by experts who have the appropriate background [8]. The assessment of content validity can be achieved through the application of multifaceted Rasch model analysis, which involves the measurement of the Rasch model [9]. Content validation is performed by expert judgment using a validation sheet containing evaluation criteria. The aspects of assessment that will be evaluated by experts through expert judgment on the SRL questionnaire instrument are (1) the relevance aspect of questionnaire statement preparation with the supporting theory used; (2) the construction and systematic aspects of questionnaire statement writing; (3) the accuracy aspect of questionnaire statements with phases and indicators of SRL; (4) the suitability aspect of language writing; and (5) the gender bias aspect.

The purpose of this study is to evaluate and analyze the quality of SRL questionnaires based on criteria developed according to non-test instrument development standards using the Many-Facet Rasch Model (MFRM) analysis. The three research questions we developed were as follows: (1) How reliable and valid are the constructs and criteria of the SRL questionnaire according to the Rasch parameters? (2) What can be learned about the criteria and questionnaire of the SRL assessment from the measurement report? (3) To what extent do raters weigh severity and leniency when assessing the quality of the SRL questionnaire?

2 Methodology

This research is a quantitative study centered on the assessment and analysis of the constructs employed in crafting SRL questionnaires, as per the predefined evaluation criteria. The evaluation was conducted by a panel of ten raters, all of whom are experts with educational backgrounds in mathematics education and educational research and assessment. Each rater is identified by a unique code, namely A, B, C, D, E, F, G, H, I, and J. The demographic data of the raters on the quality of the SRL questionnaire are presented in Table 1 below.

The evaluation tool employed has been tailored to align with the established criteria. The scoring guide for these criteria employs a four-point rating scale, with values ranging from 1 (indicating very little) to 4 (denoting adequate). These pre-defined criteria will be applied to assess the entirety of the SRL questionnaire. The questionnaire comprises 50 affirmative

statements that correspond to the 6 Zimmerman subscales. In each construct of the assessment instrument, five specific criteria are utilized for evaluation. These criteria encompass the relevance of the questionnaire statement to the underlying theory, the construction and methodical formulation of the questionnaire statement, the alignment of the questionnaire statement with the phases and SRL indicators, the appropriateness of language used in the statement, and the consideration of gender bias. A total of 10 raters will participate in the assessment process. Therefore, the complete dataset to be gathered equals 6 subscales multiplied by 5 criteria and then by 10 raters, resulting in a total of 300 data points. It's important to note that there are no missing data in this study.

Table 1. The demographic characteristics of the experts participating in this study

	Demographic	Frequency	Percentage (%)
Gender	Male	6	60%
	Female	4	40%
Age	Below 40 years old	4	40%
	40-45 years old	4	40%
	46-50 years old	2	20%
Major	Mathematics Education	6	60%
	Research and Evaluation Education	4	40%
Works Experience	Below 15 years	4	40%
	15-20 years	3	30%
	20-25 years	2	20%
	26-30 years	1	10%

The acquired data is subsequently subjected to analysis using the Many Facets Rasch Model (MFRM). MFRM is an advancement of the Rasch Model and is employed specifically for multi-rater analysis [10]. Its purpose is to accommodate and address the variability that arises from the involvement of multiple raters in the assessment process [11], [12]. One of the benefits of employing the MFRM analysis is that it allows for the individual modeling of each rater's utilization of the rating scale. As a result, the model doesn't presume that all raters will provide identical ratings [13], [14].

3 Result and Discussion

The outcomes of MFRM analysis commenced with an examination of the reliability measures for the six subscales within the SRL Questionnaire. Additionally, the analysis encompassed the five criteria employed to evaluate the quality of each subscale of the questionnaire, all of which are detailed in Table 2 below.

Table 2. The examination of reliability in the context of MFRM analysis covered both the 30-item criteria and the SRL questionnaire

	Logit Mean	Standard Deviation	Separation Index	Reliability	Standard Error
Criteria	0,00	1,32	4,79	0,96	0,04
Sub-Scale	0,00	0,71	2,26	0,84	0,01

The results, as shown in Table 2 above, indicate that both the criteria and the constructs have very good measurement reliability [15]. The criterion reliability is recorded at 0.96 (falling within the "good" category), while the subscale reliability is marked at 0.84 (also classified as "good"). In this research, the criteria yield a separation index of 4.70. This result signifies that the questionnaire's quality is strong and is capable of discerning the quality of the subscale within the SRL Questionnaire. Likewise, the subscale exhibits a separation index of 2.26, indicating that the evaluation rubric is sufficiently effective in assessing the subscale found in the SRL Questionnaire, falling within the "fairly good" category [16].

The analysis proceeded to examine the fit sequence of the statements within each criterion included in the rubric for evaluating the subscales of the SRL Questionnaire. To assess the order of item fit within the assessment criteria for each subscale of the SRL Questionnaire, the Winstep application was employed. The assessment criteria were based on the accepted mean square (MNSQ) value falling within the range of 0.5 to 1.5, the accepted z-standard (ZSTD) value ranging from -2 to +2, and the accepted point measure correlation (Pt.Mean Corr) value falling within the range of 0.4 to 0.85 [16]. The findings from the analysis of the item fit order demonstrate that all criterion items successfully pass all three assessments. Each criterion item exhibits an adequate level of quality for evaluating the quality of each subscale within the SRL Questionnaire. Moreover, the outcomes of the MFRM analysis on all criterion items reveal that these items demonstrate satisfactory unidimensionality as assessed by variance. The variance found in this research is 37.32%, meaning that each of the criteria used in the scoring rubric measures a single dimension.

Table 3 presents an overview of the logit measure quality for each subscale within the SRL Questionnaire. This assessment is conducted using the scoring rubric aligned with the compiled criteria.

Table 3. SRL questionnaire measurement report

Sub-Scale	Total Score	Infit MNSQ	Outfit MNSQ	Infit ZSTD	Outfit ZSTD
Planning	167	1,10	1,05	0,50	0,20
Self-Efficacy	176	0,86	0,74	-0,70	-0,80
Self-Satisfaction	157	0,89	0,89	-0,40	-0,40
Self-Evaluation	149	0,88	0,90	-0,50	-0,40
Self-Monitoring	156	0,94	0,88	-0,20	-0,40
Self-Efforts	163	1,32	1,37	1,40	1,40

Table 3 illustrates that every item within the SRL subscale of the questionnaire, as assessed by the experts (raters), conforms to the valid criteria employed in the MFRM analysis [9], [12]. The outcomes of the validation analysis reveal that the reliability value for all items within the SRL questionnaire subscale stands at 0.84, signifying an "excellent" rating. Furthermore, Table 3 shows that the total score that all raters assigned to each item in the evaluation criteria is quite close to the highest possible total score for all items in the SRL questionnaire subscale.

The MFRM analysis includes a Wright Map graph that shows the quality of each subscale of the SRL Questionnaire as well as the quality of the scoring criteria associated with each

subscale. It also shows the rater severity produced by the FACETS program. Figure 1 below shows the Wright map of the quality rating for the SRL subscale items.

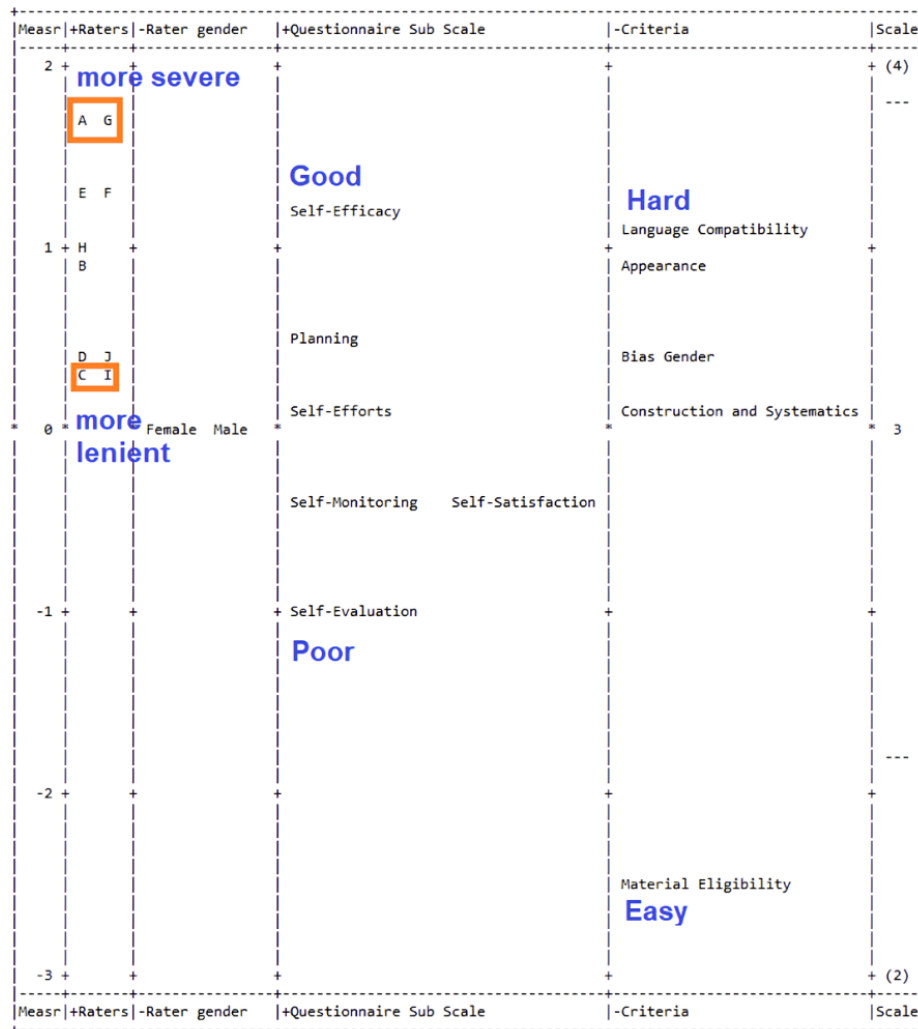


Fig. 1. Quality of the SRL questionnaire, scoring criteria, and severity distribution of the raters

From Figure 1, we can see that scorers A and G are more strict than the other scorers. Raters C and I are more lenient raters compared to the other raters. The degree to which raters applied severity to each subscale of the SRL Questionnaire instrument affects the overall quality rating of the instrument. Another finding can be seen in the fourth column, where the statements in the self-efficacy subscale are SRL subscales that are rated well by the raters. The self-evaluation subscale is the subscale in which the statements are rated least favorably by the raters. This is because the statements in the self-evaluation subscale have ambiguous meanings and overlap with statements in other subscales. Another thing is also seen in the Self

Satisfaction subscale, there are several statements that have meanings that overlap with statements in other subscales.

Based on the results of the expert validation, the experts recommended making revisions related to these statements or eliminating statements on the subscale that have ambiguous meanings. The experts also recommended reducing the number of statements because 50 statements is quite a lot. The experts felt that the large number of statements would make it uncomfortable for someone (in this case, students) to provide answers that match what they're feeling and experiencing [17]. The number of statements should be about 25-30 statements in a questionnaire [18].

Based on the results of the content validation analysis, it can be concluded that all subscales of the SRL questionnaire meet the valid criteria based on the MFRM analysis. The subscales of the SRL Questionnaire instrument are also considered as a whole to be used in the construct validation test to strengthen the results of content validation.

4 Conclusion

The acceptance of the questionnaire criteria is taken into account when evaluating the quality of the SRL Questionnaire. These standards are established by analyzing the standardized benchmarks used by the expert raters of the study. In order to analyze the quality of the questionnaire, subscales, criteria, and the degree of stringency or difficulty of raters' judgments all at once, the MFRM analysis was used. The conclusions of the analysis show that the SRL questionnaire has a good quality. In addition, the logit values obtained by the FACETS program show that the self-efficacy subscale of the SRL questionnaire has a high quality distribution in terms of quality. The linguistic appropriateness criterion is the most difficult for experts to evaluate and score. Overall, the MFRM analysis is a useful tool for evaluating the caliber of self-report learning questionnaires created using different scoring methods. In addition, it provides accurate approximations of legitimate and trustworthy metrics for rater ratings, stringency, item quality within scoring criteria, test instrument quality, and evidence of rater bias.

References

- [1] B. I. Ansari, M. Saleh, Nurhaidah, and Taufiq, "Exploring students' learning strategies and self-regulated learning in solving mathematical higher-order thinking problems," *Eur. J. Educ. Res.*, vol. 10, no. 2, pp. 743–756, 2021, doi: 10.12973/eu-jer.10.2.743.
- [2] S. M. Harding *et al.*, "Self-regulated learning as a predictor of mathematics and reading performance: A picture of students in Grades 5 to 8:," *Aust. J. Educ.*, vol. 63, no. 1, pp. 74–97, Apr. 2019, doi: 10.1177/0004944119830153.
- [3] Y. Wang and R. A. Sperling, "Characteristics of Effective Self-Regulated Learning Interventions in Mathematics Classrooms: A Systematic Review," *Front. Educ.*, vol. 5, p. 58, May 2020, doi: 10.3389/FEDUC.2020.00058/BIBTEX.
- [4] E. Araka, E. Maina, R. Gitonga, and R. Oboko, "Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—systematic review (2008–2018)," *Res. Pract. Technol. Enhanc. Learn.*, vol. 15, no. 6, 2020, doi:

10.1186/s41039-020-00129-5.

- [5] D. Bylieva, J.-C. Hong, V. Lobatyuk, and T. Nam, “Self-Regulation in E-Learning Environment,” *Educ. Sci.*, vol. 11, p. 785, 2021, doi: 10.3390/educsci11120785.
- [6] F. Gabriel, S. Buckley, and A. Barthakur, “The impact of mathematics anxiety on self-regulated learning and mathematical literacy,” *Aust. J. Educ.*, vol. 64, no. 3, pp. 227–242, 2020, doi: 10.1177/0004944120947881.
- [7] T. Toering, M. T. Elferink-Gemser, L. Jonker, M. J. G. van Heuvelen, and C. Visscher, “Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS),” *Int. J. Sport Exerc. Psychol.*, vol. 10, no. 1, pp. 24–38, Mar. 2012, doi: 10.1080/1612197X.2012.645132.
- [8] J. Van Den Akker, K. Gravemeijer, S. McKenney, and N. Nieveen, “Introducing educational design research,” in *Educational Design Research*, 4 Edition., J. Van Den Akker, K. Gravemeijer, S. McKenney, and N. Nieveen, Eds. New York: Routledge, 2006.
- [9] T. Eckes, *Introduction to many-facet Rasch measurement*. Berlin, Germany: Peter Lang Verla, 2011.
- [10] K. Kudiya, B. Sumintono, S. Sabana, and A. Sachari, “Batik Artisans’ Judgment of Batik Wax Quality and Its Criteria: An Application of the Many-Facets Rasch Model,” in *Pacific Rim Objective Measurement Symposium (PROMS)*, 2018, pp. 27–37, doi: 10.1007/978-981-10-8138-5_3.
- [11] T. G. Bond and C. M. Fox, *Applying in the Rasch Model, Fundamental Measurement in the Human Sciences*, 3rd editio. New York: Routledge, 2015.
- [12] T. Eckes, *Introduction to Many-Facet Rasch Measurement: Analyzing and Evaluating Rater-Mediated Assessments*, 2nd Editio. Berlin, Germany: Peter Lang Verla, 2015.
- [13] J. M. Linacre, “Facets computer program for many-facet Rasch measurement[Computer Software].” Winsteps.com, Beaverton, Oregon, 2022, [Online]. Available: <https://www.winsteps.com/>.
- [14] Maryati, Z. K. Prasetyo, I. Wilujeng, and B. Sumintono, “Measuring Teachers’ Pedagogical Content Knowledge Using Many-Facet Rasch Model,” *J. Cakrawala Pendidik.*, vol. 38, no. 3, pp. 452–464, Oct. 2019, doi: 10.21831/cp.v38i3.26598.
- [15] J. M. Linacre, *FACETS Version 3.71.3 [Computer Software and manual]*. Chicago: Winsteps.com, 2013.
- [16] B. Sumintono and W. Widhiarso, *Aplikasi Pemodelan Rasch pada Assessment Pendidikan [Application of Rasch Modeling to Educational Assessment]*. Cimahi: Trim Komunikata, 2015.
- [17] R. G. Kost and J. Correa Da Rosa, “Impact of survey length and compensation on validity, reliability, and sample characteristics for Ultrashort-, Short-, and Long-Research Participant Perception Surveys,” *J. Clin. Transl. Sci.*, vol. 2, pp. 31–37, 2018, doi: 10.1017/cts.2018.18.
- [18] H. Sharma, “How short or long should be a questionnaire for any research? Researchers dilemma in deciding the appropriate questionnaire length,” *Saudi J. Anesth.*, vol. 16, no. 1, pp. 65–68, 2022, doi: 10.4103/sja.sja_163_21.