Daengku: Journal of Humanities and Social Sciences Innovation *ISSN: 2775-6165 (online)*
Vol. 6 No. 3 (2026)
https://doi.org/10.35877/454RI.daengku4950

Development and Validity Based on Rush Model of Self-Regulated Scale on Gifted

Fany Rifqoh*, Muhammad Nur Wangid, Suwarjo, & Yulia Ayriza

Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

Abstract

Self-Regulated Learning (SRL) is an essential skill that enables learners to independently plan, monitor, and evaluate their learning processes. Despite the availability of several SRL scales, valid and context-appropriate instruments tailored to the characteristics of gifted students in madrasah accelerated programs remain scarce. This study aims to examine the validity of a Self-Regulated Learning (SRL) instrument adapted specifically for gifted students enrolled in accelerated programs within Indonesian madrasahs. The research utilizes an adapted version of the Erdogan SRL Scale, integrating components from the Motivated Strategies for Learning Questionnaire (MSLQ) and the Self-Regulatory Learning Interview Schedule (SRLIS). A total of 361 gifted students from various regions in Indonesia participated. Data were analyzed using the Rasch model assisted by R software to evaluate item fit, reliability, and threshold performance. A total of 67 items were confirmed to meet the Rasch model eligibility criteria, with no indications of overfit or misfit. Reliability values for both items and participants were within acceptable ranges. However, several items showed suboptimal thresholds and were recommended for revision to enhance clarity and discriminatory power. The adapted SSRL instrument is valid for measuring the SRL abilities of gifted students in madrasah accelerated programs. The findings support its use in educational assessment and intervention design, offering educators a reliable tool to better understand and strengthen self-regulatory learning skills among gifted learners.

Keywords: Acceleration Program, Independent Learning, Student Superior Ability, Rasch Model.

Received: 3 March 2026 Revised: 28 May 2026 Published: 30 April 2026

1. Introduction

Self-Regulated Learning (SRL) is known as a learning approach that emphasizes the ability of students to independently regulate, monitor, and evaluate their learning process (Xu et al. 2023). In the context of learning in madrasahs, this ability is related to students' capacity to maintain excellent performance through effective self-management (Ridgley, DaVia Rubenstein, & Callan, 2022; Butler, 2023; Siekanska, Wilson, Blecharz, & Young, 2023). A number of studies show that academic success is generally possessed by students who are confident, persistent, and have mature learning strategies (Fan et al. 2025). They view the learning process as a purposeful activity and take full responsibility for achieving academic goals (Banihashem et al. 2025). Thus, SRL occupies an important position for gifted students (Ridgley et al., 2022), especially when they are faced with complex tasks and a learning environment that provides a high degree of autonomy (Hertel, A. & B., 2024).

One form of educational service for gifted students in madrasahs is an acceleration program. To participate optimally in this program, students need strong self-management skills, including SRL (Zimmerman, Bonner, & Kovach, 1996; Wangid, 2004; Wahab, 2010; Nilson, 2013; Hertel, 2024). Therefore, mapping SRL levels is an important basis for designing more appropriate learning interventions. This effort requires the availability of measurement tools that are appropriate for the characteristics of gifted students in accelerated programs, but valid and relevant instruments are still limited.

Early research on SRL was conducted by Zimmerman (2023), which produced an interview-based tool. Although informative, the validity of this instrument has not been reported, and it is not efficient for large-scale measurement. Another instrument was developed by Marougkas et al. (2024) through the MSLQ, which has been validated and widely

* Corresponding author.
E-mail address: fanyrifqoh.2023@student.uny.ac.id

used to measure motivation and self-reported learning strategies. However, the MSLQ has limitations in terms of self-report bias, fails to capture the dynamics of the learning process, and requires cultural adaptation to be context-appropriate. Meanwhile, the instrument developed by Karaman (2024) is still limited to the Turkish context and is not easily accessible, requiring flexibility for wider use.

In addition, Hakim (2024) notes that failure to achieve learning targets can trigger stress, one of which is through academic anxiety, and this phenomenon is also commonly found in gifted students. Based on interviews with guidance counselors, a number of students with superior abilities actually participate in remediation and report anxiety more often than low effort. This dimension is also reflected in the scale developed by Erdogan. Research by Larasati, Supahar, and Yunanta (2020) states that of the 11 SRL instruments that have been developed previously, the measurement of high-ability students is still not in-depth. This gap reinforces the need for culturally relevant instruments that are appropriate for gifted students in Indonesia, especially those who are enrolled in accelerated programs in madrasahs.

Overall, this study emphasizes the importance of developing and validating SRL scales that are appropriate for the needs of gifted students. Such instruments are essential not only for assessment purposes but also to support counseling and learning interventions. The limitations of construct-validated instruments are further emphasized in recent literature (Qin et al., 2024; Carter et al., 2020; Ridgley et al., 2022).

Measuring instruments that are not tailored to the characteristics of the target population or are not well standardized can produce biased data and interfere with the decision-making process. Therefore, the development of an SRL scale specifically designed for high-achieving students in madrasahs, and tested with a robust methodological approach, is an urgent scientific and practical need. In line with this objective, this study focuses on the development of an SRL instrument and testing its validity to support the optimization of high-achieving students' abilities in madrasahs.

This study has unique characteristics through the development of a Self-Regulated Learning (SRL) scale specifically for high-achieving students in accelerated programs in madrasahs. This scale is a contextual adaptation of Erdogan's SSRL (2016), taking into account aspects of language, culture, and the characteristics of Islamic education services that are distinctive to madrasah.

The uniqueness of this study also lies in the use of the Rasch Model-based validation method as a modern approach in instrument development, which allows for a more comprehensive evaluation of item accuracy, participant response quality, and model suitability. The resulting instrument is expected to be valid, reliable, and culturally and pedagogically relevant, so that it can be used both as an assessment tool and as a basis for designing independent learning interventions for gifted students in madrasahs.

In the process of testing construct validity, the selection of appropriate statistical software is an important aspect to ensure accurate analysis results (Haerudin & Iyan Rosita Dewi Nur, 2020). R Studio was used to support Rasch model-based analysis, which is considered a cutting-edge approach in developing the SRL scale adapted from Erdogan's SSRL. The developed instrument was then implemented on a sample of high-ability students participating in an acceleration program in Indonesia as part of the validation process.

2. Literature Review

2.1. Student Acceleration & Superior Ability Program

An acceleration program is a form of special educational service for students with potential intelligence or special talents, where they complete the curriculum faster than regular students. According to Nawawi and Swandari (2022), accelerated programs in madrassas involve not only rigorous selection but also adjustments to curriculum and learning methods to stimulate higher-level thinking, such as through PAKEM, team learning, modules, and psychological management. Astuti, Hanafi, & Sarkadi (2022) examined the management of accelerated programs for gifted children in public schools, and found that although the implementation is quite good, strong program supervision is still needed from school principals, education offices, and parents to maintain the stability of students' academic performance.

2.2. Self-Paced Learning

Self-directed learning is becoming increasingly important, especially in the context of online learning. Subekti (2020) found that in online English classes (EAP), students' self-paced learning orientation was very weakly associated with academic achievement, highlighting the challenges of providing effective self-support in an online setting. In addition, Rahmawati & Irawan (2024) examined how learning anxiety impacts the mathematics achievement of elementary school students and found that self-directed learning contributes significantly positively to achievement, suggesting that building learning independence can be an important strategy to improve learning outcomes.

2.3. Model Rasch

The Rasch model has been widely used in the validation of SRL (Self-Regulated Learning) instruments. For example, Ramadhani, Syahputra, & Simamora (2024) used Rasch's analysis on an SRL questionnaire (50 items) for high school students and found that 25 statements met the criteria of construct validity, gender inclusivity, and unidimensionality. Nizaruddin et al. (2024) also evaluated the SRL scale with the Rasch approach in mathematics education students. From the initial 30 items, the scale was refined to 28 items after three calibrations and showed good reliability and validity.

In the physics subject research, Irsalina, Aviyanti, & Rahayani (2025) examined the SRL profile of grade 11 students using the Rasch and MSLQ (Motivated Strategies for Learning Questionnaire) models. They found that most students had a moderate SRL profile, but aspects such as test anxiety, critical thinking, and effort regulation were still weak. In addition, Mumpuni, Begimbetova, & Retnawati (2023) conducted a DIF (Differential Item Functioning) analysis on the SRL questionnaire with Rasch and detected several mis-fit items as well as bias related to the length of the study, although the total reliability was very high.

Finally, Dewanti, Izzah, & Kiranasari (2024) used IRT (Item Response Theory) with the Partial Credit model (Rasch variant) to test self-regulation instruments in mathematics learning. They found five important aspects: self-confidence, discipline, active learning, responsibility, and motivation.

3. Methods

3.1. Participants

In this study, respondents were recruited from various madrasahs that run accelerated programs. The first permit was sent to the Regional Office of the Ministry of Religious Affairs in Yogyakarta Province because Yogyakarta Province had the largest sample size. After obtaining permission from the province, the researchers were able to provide copies of the permit to the targeted madrasahs in Yogyakarta. For East Java Province, the research permit process was carried out directly at the madrasahs. The researchers provided two alternatives for administering the questionnaire: asking students to fill out the scale directly or through Google Forms. This study used 361 gifted students participating in the acceleration program at the MTs level in Java as respondents. Based on the above data, according to Joreskog (1968), it is recommended that the sample size be at least 5 (five) times the number of parameters in the CFA model if the estimation method used is “maximum likelihood,” or 10 times if using another estimator. The number of items in the Self-Regulation-Based Learning Scale developed by the researcher is 67 items, so the sample size meets the requirements.

3.2. Intruments

The SRL instrument developed in this study uses a five-point Likert scale (1 = Never to 5 = Always). This instrument was adapted from the Self-Regulation in Learning Scale (SSRL) by Tolga Erdogan (2016), which integrates elements from Pintrich & De Groot's MSLQ and Zimmerman & Martinez-Pons' SRLIS. Erdogan's SSRL is a multidimensional construct consisting of 17 dimensions and 67 items, which include self-learning skills and motivation. Cognitive aspects are divided into three phases: before learning (e.g., environmental structuring, planning), during learning (e.g., organizing information, seeking help, self-monitoring), and after learning (e.g., self-evaluation, consequences after success or failure). Motivational aspects include task value, self-efficacy, anxiety, failure attribution, and goal orientation.

Formal permission to use Erdogan's scale was obtained through a letter of request, to which Erdogan responded positively, providing three versions of the instrument: an initial draft, a validated version, and one version categorized by item polarity. Adaptation steps included translation and contextualizing the items to suit gifted learners in Indonesia, with the help of expert judgment. Instrument grid shown on Table 1.

3.3. Data Collection

After developing the 67 items, the instrument was reviewed by three experts-one linguist and two instrument experts-who evaluated the relevance of each item to measure self-regulated learning in high-ability students using a four-point scale. They also provided qualitative feedback for improvement. Quantitative analysis using Gregory's formula was conducted to assess content validity, while qualitative feedback helped to improve item clarity and grammar. After revision, the final instrument consisted of 67 items for the pilot study.

The pilot study used both offline (paper-based) and online (Google Form) methods. Offline data collection involved coordination with schools and classroom-based distribution, where students completed the questionnaire within 20-25 minutes. Online data collection took place over three weeks in November 2024 using a dedicated link. Participants could only respond once and were informed of confidentiality and voluntary participation to ensure ethical standards were upheld.

Table 1. Instrument Grid

Steps	Original Item	Adjustments for Gifted-Accelerated Students
Before Study	I find it difficult to decide how to use my time effectively. I consult with my teacher on subjects that are important for the long term. At the beginning of the semester, I plan what to do and when to do it.	I sometimes have difficulty dividing my study time because the accelerated material requires faster mastery. I consulted with my teacher about subjects that are important for the long term/during the accelerated junior high school period. I designed a semesterly study strategy, so that I could complete the material more quickly and in-depth at my own pace.
During Study	I only read reference books given by the teacher.	I only read the PKBM book (Independent Learning Activity Guide) as a reference given by the teacher.
Motivation	I am confident that I can understand even the most difficult subject matter.	I believe I can understand the challenging accelerated materials with the right effort and learning strategies.

3.4. Data Analysis

Data analysis in this study was conducted in two stages. The first stage used descriptive statistics to assess data feasibility and describe the basic characteristics of respondents before further analysis was conducted. The second stage utilized Rasch Model-based analysis to test the validity of self-regulated learning constructs. The Rasch model was used to accurately assess item suitability, participant response quality, and item difficulty structure to ensure that the developed instrument adequately represented the construct.

In Rasch analysis, items are considered to have an ideal level of difficulty if they fall within the range of -2 to 2 (Aryadoust et al. 2021). In addition, response category thresholds must show a monotonically increasing order ($b1 < b2 < b3 < b4$), with relatively proportional distances between thresholds (Andrich, 1978; Aryadoust et al. 2021). Infit statistics are used to identify potential item mismatches directly related to participant response patterns. Infit and outfit values are considered optimal in the range of 0.7 to 1.3, although a range of 0.5 to 1.5 is still acceptable for instrument interpretation (Aryadoust et al. 2021; Bond et al. 2020).

With this approach, the Rasch Model allows for a more in-depth assessment of item accuracy and measurement structure consistency, so that construct validity results can be used as a strong basis for developing an SRL scale for gifted students in madrasahs.

4. Result and Discussion

4.1. Descriptive Statistics

The self-regulated learning scale consists of two subscales, namely independent learning skills and motivation. Self-regulated learning skills (cognitive factors) are grouped into three main dimensions, namely before study (organizing/managing learning time, planning, and managing the environment), during study (organizing and transforming, searching for appropriate information, searching for easily accessible information, seeking help from peers, teachers, or adults, self-monitoring and practicing and memorizing), and after study (self-evaluation, self-consequences after success, and self-consequences after failure). Motivational factors include task value, self-efficacy, anxiety, failure attribution, and goal orientation. There are a total of 17 subdimensions consisting of 67 statement items.

4.2. Validity Based on Rasch Model

To further validate the construction of the self-regulated learning scale, with Rasch model analysis, measured through two subscales and 17 dimensions, with the empirical data.

4.3. Rasch Model Analysis on SMB Dimensions

In Table 1, items on the SMB dimension have difficulty parameters ranging from -1.18 to 0.25 (M = -0.51, SD = 0.47). This indicates that the items have ideal difficulty parameters as they are in the range of -2 to 2 (Aryadoust et al. 2021). Item B7 is the item where it is most difficult for respondents to select the "always" category (score 5), while item B12 is the item where it is easiest for respondents to select the "never" category (score 1) on the instrument. The threshold parameter is the boundary between two consecutive categories on the Likert scale used, where when the scale consists of 5 categories then the item will have four thresholds (i.e., b1, b2, b3, and b4). Thresholds should be in monotonically ascending order, such that b1 < b2 < b3 < b4, where the distance between thresholds is suggested to be fairly even (Andrich, 1978; Aryadoust et al. 2021). Table 1 shows that there are three items, B1, B3 and B5 with non-ideal thresholds. Item B1 shows b2 > b3, indicating that the ability required to select the "sometimes" category (score 3) is higher than the ability to select the "usually" category (score 4). Item B3 shows b2 = b3, indicating that the ability required to select the "sometimes" category (score 3) is the same as the ability to select the "usually" category (score 4). Item B5 shows that b2 > b3, indicating that the ability required to select the "sometimes" category (score 3) is higher than the ability to select the "usually" category (score 4). Items with non-ideal thresholds were considered for revision.

Table 2. Item Parameters on the SMB Dimension

Item	Difficulty	Threshold			
		b1	b2	b3	b4
B1	-0.25	-1.16	-0.04	-0.30	0.49
B2	-0.70	-2.76	-0.53	-0.32	0.79
B3	-0.82	-1.59	-0.82	-0.82	-0.06
B4	-0.08	-1.73	-0.42	0.61	1.21
B5	0.24	-0.74	0.22	0.20	1.28
B6	-0.60	-2.02	-0.82	-0.37	0.80
B7	0.25	-0.24	0.31	0.41	0.53
B8	-0.43	-1.98	-0.49	-0.08	0.85
B9	-0.35	-1.09	-0.81	0.05	0.42
B10	-1.10	-1.96	-0.96	-0.91	-0.56
B11	-0.96	-2.10	-1.25	-0.64	0.14
B12	-1.18	-1.85	-1.45	-0.82	-0.60
B13	-0.69	-0.85	-0.84	-0.80	-0.27

Table 2 presents the item fit statistics on the SMB dimension. Outfit statistics are used to detect unexpected responses (outliers) in very easy or very difficult items. Meanwhile, infit statistics are used to detect mismatches in items that are relevant to the respondents. The ideal infit and outfit statistics range from 0.7 to 1.3, although values ranging from 0.5 to 1.5 remain acceptable (Aryadoust et al. 2021; Bond et al. 2020). On the SMB dimension, items had outfit statistics ranging from 0.83 to 1.14 (M = 0.93, SD = 0.11), while infit statistics ranged from 0.84 to 1.11 (M = 0.94, SD = 0.09). This indicates that all items were neither overfit (data too idealized) nor misfit (random or unmodeled response patterns).

Figure 1 is a Wright Map illustrating the correspondence between respondents' abilities and the four thresholds on the SMB dimension. Figure 1 indicates that the distribution of respondents' abilities looks quite symmetrical and is around 0. Threshold items are spread from around -3 to 1. However, some b4 thresholds are higher than 0, and some b1 thresholds are very low. This indicates that some of the extreme categories (i.e. "always" or "never") may have been chosen infrequently or were too far-reaching for respondents of average ability. In Figure 1, not all thresholds show the ideal order (i.e. B1, B3 and B5), indicating that the choice of response categories is not functioning properly (category dysfunction). In addition, there are some items with overlapping thresholds, such as Items B10 and B13, indicating that respondents' responses on these items do not distinguish much between ability levels. Items identified with category dysfunction and overlapping were considered for revision.

Table 3. Item Fit Statistics on SMB Dimensions

Item	Outfit	z-Outfit	Infit	z-Infit
B1	1.144	2.383	1.112	1.915
B2	0.887	-1.914	0.884	-1.973
B3	1.057	0.727	1.028	0.387
B4	1.113	1.650	1.110	1.599
B5	0.957	-0.714	0.964	-0.597
B6	0.885	-1.764	0.886	-1.760
B7	0.914	-1.407	0.910	-1.538
B8	0.863	-2.310	0.864	-2.296
B9	0.829	-2.829	0.838	-2.682
B10	0=47.836	-1.867	0.902	-1.145
B11	0.860	-1.983	0.877	-1.757
B12	0.835	-1.886	0.896	-1.199
B13	0.938	-0.703	0.971	-0.334

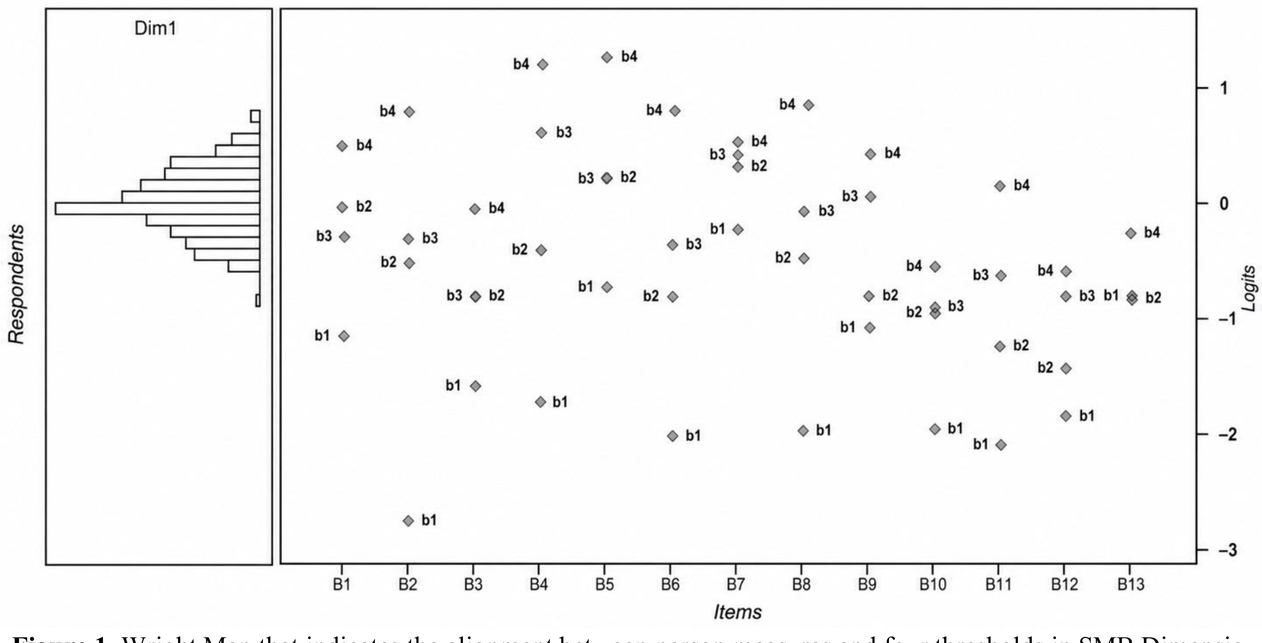

Figure 1. Wright Map that indicates the alignment between person measures and four thresholds in SMB Dimensio.

4.4. Rasch Model Analysis on SB Dimension

In Table 3, items in the SB dimension have difficulty parameters ranging from -1.34 to 0.56 (M = -0.47, SD = 0.54). This can be interpreted that the items have ideal difficulty parameters because they are in the range of -2 to 2 (Aryadoust et al. 2021). Item B19 is the item where it is most difficult for respondents to select the "always" category (score 5), while item B31 is the item where it is easiest for respondents to select the "never" category (score 1) on the instrument.

The threshold parameter is the boundary between two consecutive categories on the Likert scale used, where when the scale consists of 5 categories then the item will have four thresholds (i.e., b1, b2, b3, and b4). Thresholds should be in monotonically ascending order, such that b1 < b2 < b3 < b4, where the distance between thresholds is suggested to be fairly even (Andrich, 1978; Aryadoust et al. 2021). Table 1 shows that there are three items, namely B19, B30, and B31 with non-ideal thresholds. Item B19 shows b2 > b3, indicating that the ability required to select the "sometimes" category (score 3) is higher than the ability to select the "usually" category (score 4). Item B30 shows b2 > b3, indicating that the ability required to select the "sometimes" category (score 3) is higher than the ability to select the "usually" category (score 4). Item B31 shows b2 > b3, indicating that the ability required to select the "sometimes" category (score 3) is higher than the ability to select the "usually" category (score 4). Items with non-ideal thresholds are considered for revision.

The item fit statistics on the SB dimension. Outfit statistics are used to detect unexpected responses (outliers) in very easy or very difficult items. Meanwhile, infit statistics are used to detect mismatches in items that are relevant to the respondents. Ideal infit and outfit statistics range from 0.7 to 1.3, although values ranging from 0.5 to 1.5 are acceptable (Aryadoust et al. 2021; Bond et al. 2020). On the SB dimension, items had outfit statistics ranging from 0.83 to 1.21 (M = 0.96, SD = 0.12), while infit statistics ranged from 0.83 to 1.19 (M = 0.95, SD = 0.1). This indicates that all items were neither overfit (data too idealized) nor misfit (random or unmodeled response patterns). Ideal MNSQ criteria (Aryadoust et al. 2021; Bond et al. 2020).

Wright Map illustrating the correspondence between respondents' abilities and the four thresholds on the SB dimension. Figure 2 indicates that the distribution of respondents' abilities looks quite symmetrical and is around 0. Threshold items are spread from around -3 to 2. However, some b4 thresholds are higher than 0, and some b1 thresholds are very low. This indicates that some of the extreme categories (i.e. "always" or "never") may have been chosen infrequently or were too far-reaching for respondents of average ability. In Figure 2, not all thresholds show the ideal order (i.e.: B19, B30, and B31), indicating that the choice of response categories is not functioning properly (category dysfunction). In addition, there are some items with overlapping thresholds, such as Items B19, B23 and B32, indicating that respondents' responses on these items do not distinguish between ability levels. Items identified with category dysfunction and overlapping were considered for revision.

4.5. Rasch Model Analysis on STB Dimensions

The STB dimension have difficulty parameters ranging from -1.39 to 0.69 (M = -0.22, SD = 0.68). This indicates that items have ideal difficulty parameters as they are in the range of -2 to 2 (Aryadoust et al. 2021). Item B44 is the item where it is most difficult for respondents to select the "always" category (score 5), while item B33 is the item where it is easiest for respondents to select the "never" category (score 1) on the instrument. The threshold parameter is the boundary between two consecutive categories on the Likert scale used, where when the scale consists of 5 categories then the item will have four thresholds (i.e., b1, b2, b3, and b4). Thresholds should be in monotonically ascending order, such that b1 < b2 < b3 < b4, where the distance between thresholds is suggested to be fairly even (Andrich, 1978; Aryadoust et al. 2021). Shows that there are six items, namely B36, B39, B40, B41, B42, and B45 with non-ideal thresholds. Item B36 shows b2 > b3, indicating that the ability required to select the "sometimes" category (score 3) is higher than the ability required for "usually" (score 4). Item B39 shows b1 > b2 > b3, indicating that the ability to select the "rarely" category (score 2) is higher than the ability required to select the "sometimes" category (score 3) is higher than the ability to select the "usually" category (score 4). Item B40 shows that b1 > b2, indicating that the ability required to select the "rarely" category (score 2) is higher than the ability to select the "sometimes" category (score 3). Item B41 shows b3 = b4, indicating that the ability required to select the "usually" category (score 4) is the same as the ability required to select the "always" category (score 5). Item B42 shows b3 > b4, indicating that the ability required to select the "usually" category (score 4) is higher than the ability required to select the "always" category (score 5). Item B45 shows b2 > b3, indicating that the ability required to select the "sometimes" category (score 3) is higher than the ability to select the "usually" category (score 4). Items with non-ideal thresholds are considered for revision.

The item fit statistics on the STB dimension. Outfit statistics are used to detect unexpected responses (outliers) in very easy or very difficult items. Meanwhile, infit statistics are used to detect mismatches in items that are relevant to the respondents. Ideal infit and outfit statistics range from 0.7 to 1.3, although values ranging from 0.5 to 1.5 are acceptable (Aryadoust et al. 2021; Bond et al. 2020). On the STB dimension items had outfit statistics ranging from 0.80 to 1.09 (M = 0.93, SD = 0.08), while infit statistics ranged from 0.81 to 1.06 (M = 0.94, SD = 0.07). This indicates that all items were neither overfit (data too idealized) nor misfit (random or unmodeled response patterns).

Wright Map illustrating the correspondence between respondents' abilities and the four thresholds on the SMB dimension. Figure 1 indicates that the distribution of respondents' abilities looks quite symmetrical and is around 0. Threshold items are spread from around -3 to 1. However, some b4 thresholds are higher than 0, and some b1 thresholds are very low. This indicates that some of the extreme categories (i.e. "always" or "never") may have been chosen infrequently or were too far-reaching for respondents of average ability. In Figure 3, not all thresholds show the ideal order (i.e.: B36, B39, B40, B41, B42 and B45), indicating that the choice of response categories is not functioning properly (category dysfunction). In addition, there are several items with overlapping thresholds, such as Items B35, B37, B39, B40, B44, and B45 indicating that respondents' responses on these items do not distinguish between ability levels. Items identified with category dysfunction and overlapping were considered for revision.

Table 4. Item Parameters on STB Dimensions

Item	Difficulty	Threshold			
		b1	b2	b3	b4
B33	-1.39	-3.00	-1.24	-0.99	-0.34
B34	-1.00	-2.50	-1.03	-0.57	0.08
B35	-0.74	-1.22	-1.16	-0.60	0.01
B36	0.62	-0.45	0.89	0.59	1.43
B37	0.38	-0.09	-0.03	0.48	1.15
B38	-0.19	-1.24	-0.34	0.10	0.73
B39	-0.17	-0.03	-0.16	-0.52	0.03
B40	-0.31	-0.26	-0.48	-0.44	-0.07
B41	-0.91	-1.46	-1.34	-0.41	-0.41
B42	-0.59	-1.13	-0.58	-0.25	-0.38
B43	0.28	-0.40	0.06	0.59	0.86
B44	0.69	0.42	0.44	0.75	1.16
B45	0.51	0.27	0.46	0.32	0.98

Table 5. Item Fit Stats on STB Dimensions

Item	**Outfit**	**z-Outfit**	**Infit**	**z-Infit**
B33	0.940	-0.715	0.968	-0.376
B34	0.947	-0.761	0.956	-0.638
B35	0.992	-0.076	0.985	-0.170
B36	1.087	1.131	1.061	0.830
B37	0.956	-0.746	0.960	-0.671
B38	1.031	0.526	1.027	0.469
B39	0.817	-3.410	0.818	-3.591
B40	0.800	-3.298	0.813	-3.263
B41	0.840	-2.165	0.866	-1.847
B42	0.886	-1.746	0.899	-1.616
B43	0.984	-0.244	0.989	-0.163
B44	0.932	-0.901	0.944	-0.777
B45	0.920	-1.208	0.927	-1.158

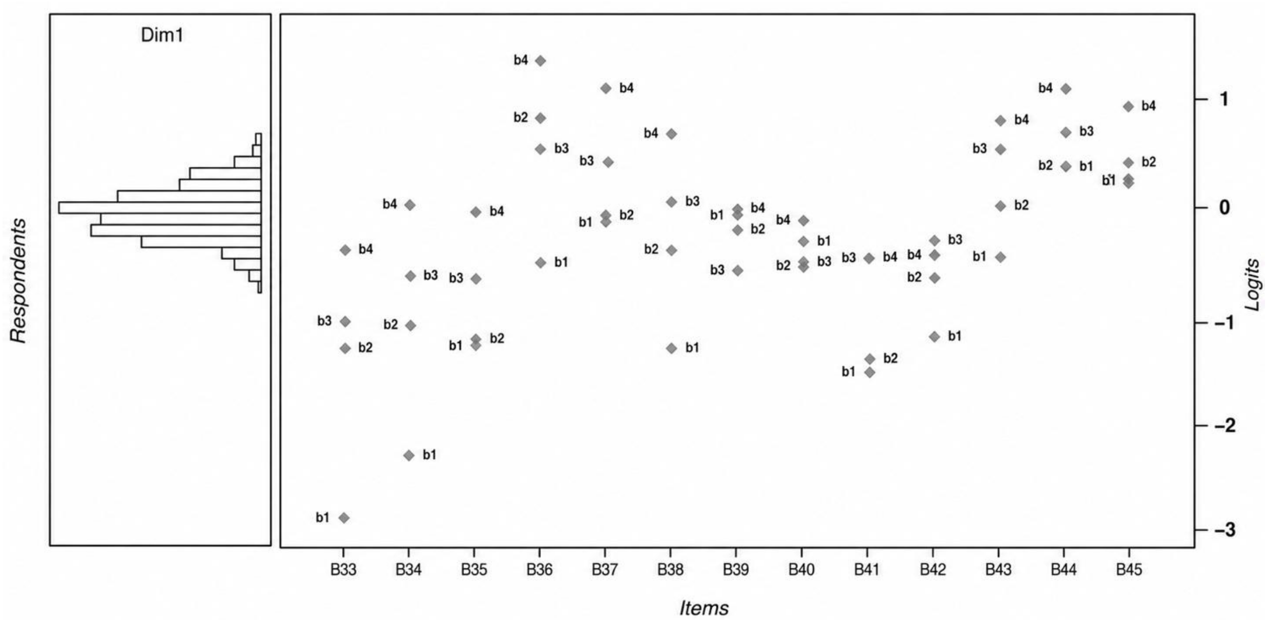

Figure 2. Wright Map that indicates the alignment between person measures and four thresholds in STB Dimension

4.6. Rasch Model Analysis on MTB Dimensions

The MTB dimension have difficulty parameters ranging from -1.44 to 0.23 (M = -0.64, SD = 0.53). This indicates that the items have ideal difficulty parameters as they are in the range of -2 to 2 (Aryadoust et al. 2021). Item B62 is the item where it is most difficult for respondents to select the "always" category (score 5), while item B49 is the item where it is easiest for respondents to select the "never" category (score 1) on the instrument. The threshold parameter is the boundary between two consecutive categories on the Likert scale used, where when the scale consists of 5 categories then the item will have four thresholds (i.e., b1, b2, b3, and b4). Thresholds should be in monotonically ascending order, such that b1 < b2 < b3 < b4, where the distance between thresholds is suggested to be fairly even (Andrich, 1978; Aryadoust et al. 2021). Table 7 shows that there are fourteen items, namely B48, B49, B51, B52, B53, B54, B55, B56, B57, B58, B61, B62, B66, and B67 with non-ideal thresholds. Item B48 shows b1 > b2, indicating that the ability required to select the "sometimes" category (score 3) is higher than the ability to select the "usually" category (score 4). Items with non-ideal thresholds were considered for revision.

The item fit statistics on the MTB dimension. Outfit statistics are used to detect unexpected responses (outliers) in very easy or very difficult items. Meanwhile, infit statistics are used to detect mismatches in items that are relevant to the respondents. Ideal infit and outfit statistics range from 0.7 to 1.3, although values ranging from 0.5 to 1.5 are acceptable (Aryadoust et al. 2021; Bond et al. 2020). On the MTB dimension, items had outfit statistics ranging from 0.87 to 1.12 (M = 0.96, SD = 0.06), while infit statistics ranged from 0.90 to 1.07 (M = 0.96, SD = 0.04). This indicates that all items were neither overfit (data too idealized) nor misfit (random or unmodeled response patterns).

Wright Map that illustrates the correspondence between respondents' abilities and the four thresholds on the MTB dimension. The distribution of respondents' abilities looks quite symmetrical and is around 0. Threshold items are spread from around -3 to 1. However, some b4 thresholds are higher than 0, and some b1 thresholds are very low. This indicates that some of the extreme categories (i.e. "always" or "never") may be rarely chosen, or too far-reaching for respondents of average ability. In Figure 4, almost all thresholds show a non-ideal order (i.e. B1, B3 and B5), indicating that the choice of response categories is not functioning properly (category dysfunction). In addition, there are some items with overlapping thresholds, such as Item B13, indicating that respondents' responses on this item do not distinguish between ability levels. Items identified with category dysfunction and overlapping were considered for revision.

Our study reveals that the SRL scale for accelerated learners is valid. Our study proved that there are four important factors to measure SRL: before study, during study, after learning, and motivation. In addition, the 17 dimensions contained in these factors made significant contributions with varying loadings. The proof of construct validity indicates that the test measures the theoretical ability or construct to be measured (Aiyegbusi et al. 2024; Singh et al. 2022;

Retnawati, 2016) and the test construction is developed based on an appropriate conceptual framework (Aiyegbusi et al. 2024; Retnawati, 2016). The findings of our study showing that there are four main components that significantly measure learners' SRL ability at the junior high school level are consistent with the findings of previous studies (Erdogan, 2016).

Table 6. Item Parameters on MTB Dimensions

Item	Difficulty	Threshold			
		b1	b2	b3	b4
B46	-0.55	-1.62	-1.10	-0.03	0.55
B47	-1.17	-2.42	-1.72	-0.72	0.18
B48	-1.02	-0.57	-2.88	-0.77	0.16
B49	-1.44	-2.17	-2.60	-0.75	-0.21
B50	-1.26	-1.60	-1.38	-1.26	-0.80
B51	-1.20	-0.76	-0.96	-1.70	-1.39
B52	-0.70	-0.60	-1.01	-0.67	-0.50
B53	-0.12	-0.13	-0.35	0.47	-0.46
B54	-1.32	-1.72	-0.91	-1.29	-1.35
B55	-0.98	-0.72	-1.65	-1.12	-0.42
B56	-0.95	-1.43	-1.43	-0.74	-0.18
B57	-1.27	-0.02	-3.03	-1.32	-0.71
B58	-0.59	-0.89	-1.59	-0.24	0.34
B59	0.04	-0.81	-0.20	0.51	0.66
B60	-0.01	-0.91	-0.26	0.40	0.74
B61	-0.28	-0.73	-1.08	0.22	0.46
B62	0.23	-0.18	0.12	0.65	0.33
B63	-0.27	-1.24	-0.20	-0.01	0.36
B64	0.17	-0.74	0.03	0.41	0.99
B65	-0.30	-0.83	-0.46	-0.12	0.19
B66	-0.69	-1.10	-0.69	-0.40	-0.56
B67	-0.39	-0.99	-0.81	0.13	0.10

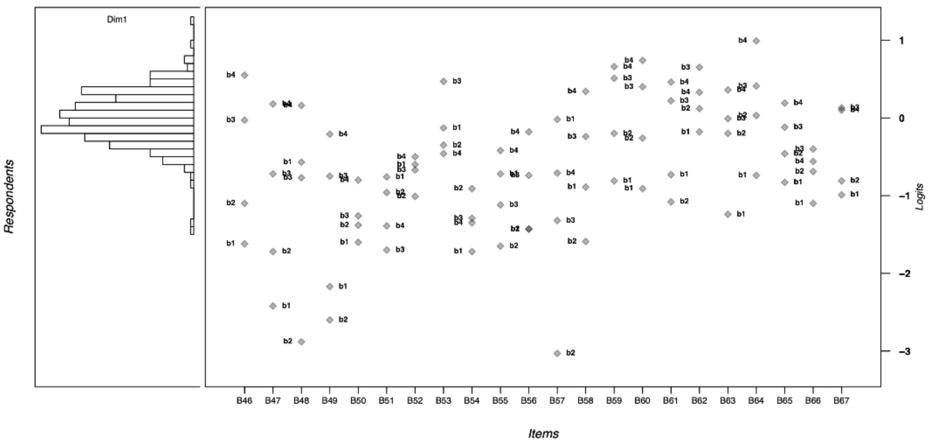

Figure 4. Wright Map that indicates the alignment between person measures and four thresholds in MTB Dimension

Table 7. Item Fit Statistics on MTB Dimensions

Item	Outfit	z-Outfit	Infit	z-Infit
B46	1.003	0.065	0.991	-0.119
B47	1.011	0.170	0.990	-0.116
B48	1.017	0.232	0.991	-0.084
B49	0.949	-0.698	0.950	-0.703
B50	0.951	-0.408	0.948	-0.463
B51	1.018	0.165	0.934	-0.388
B52	1.023	0.283	0.969	-0.360
B53	0.951	-0.856	0.952	-0.883
B54	0.871	-0.905	0.937	-0.461
B55	0.965	-0.313	0.970	-0.267
B56	0.959	-0.466	0.939	-0.743
B57	0.870	-1.206	0.913	-0.776
B58	1.029	0.410	1.014	0.218
B59	1.002	0.051	0.999	0.002
B60	0.909	-1.506	0.914	-1.432
B61	0.921	-1.197	0.922	-1.196
B62	1.116	1.791	1.073	1.194
B63	0.925	-1.289	0.918	-1.429
B64	0.901	-1.645	0.898	-1.709
B65	0.933	-1.077	0.931	-1.157
B66	0.865	-1.783	0.899	-1.449
B67	0.985	-0.212	0.974	-0.403

4.7. Discussion of Main Results

The present study examined the psychometric validity of a Self-Regulated Learning (SRL) scale specifically adapted for gifted students enrolled in accelerated programs within Indonesian madrasahs. The Rasch model analysis supported the structural integrity of the four SRL dimensions skills before learning (SMB), skills during learning (SB), skills after learning (STB), and motivational factors (MTB) indicating that these constructs function coherently within the context of high-ability learners. The distribution of item difficulty parameters, which fell within the ideal range, suggests that the items successfully captured varying levels of SRL abilities and were sensitive to differences among gifted students. This finding reinforces the understanding that gifted learners exhibit diverse patterns of cognitive preparation, learning management, and reflective evaluation, even within academically homogeneous accelerated environments.

Another noteworthy finding is the absence of overfit or misfit in the item fit statistics, which indicates that the respondents engaged with the instrument consistently and meaningfully. This strengthens the scale's construct validity and confirms that the adapted items are aligned with students' internal representations of self-regulated learning. These results harmonize with previous literature asserting that gifted students generally possess strong metacognitive awareness and advanced cognitive processing abilities that enable them to engage more effectively with structured self-evaluation tools. However, despite good overall model fit, the identification of threshold disorder across several items highlights an important nuance in how students interpret and differentiate Likert-scale categories.

The threshold irregularities particularly the overlap between "rarely," "sometimes," and "usually" imply that students may not perceive clear distinctions between the frequency descriptors as intended. This aligns with earlier studies showing that young learners, including gifted students, often rely on intuitive judgments rather than calibrated frequency estimation when responding to self-report items. These findings emphasize the importance of refining response categories or adjusting item phrasing to better accommodate students' developmental understanding of behavioral frequency.

In addition, the presence of category overlap suggests that while the cognitive dimensions of the scale were well-understood, the affective and motivational items may require greater contextual clarity. Gifted students in accelerated programs often face conflicting motivational pressures high expectations, academic competitiveness, and self-imposed performance standards which can influence their self-perceptions of motivation and effort. Therefore, items measuring motivation may benefit from contextual anchoring to elicit more accurate responses.

The results of this study also align with theoretical perspectives on gifted education which argue that self-regulated learning develops differently among high ability students. Although gifted learners are often assumed to be inherently self-regulated, previous research has shown that they may struggle with emotional regulation, persistence, or strategic planning when placed in high-demand academic environments. The current findings reinforce this view, demonstrating that SRL is not uniformly strong across all gifted students and requires systematic assessment and targeted support.

Another important implication emerging from this research is the relevance of cultural and institutional context. Madrasah accelerated programs operate with unique curricular structures, religious environments, and learning expectations that may shape SRL behaviors differently compared to general schools. The validation of the SRL scale within this specific context contributes significantly to the growing body of literature on culturally responsive assessment tools for high-ability learners.

Finally, the successful use of the Rasch model in this study highlights the importance of modern psychometrics in educational measurement. Unlike classical test theory, the Rasch model provides deeper insights into item functioning, person ability, and scale structure. This approach enables researchers and educators to identify subtle issues such as threshold disorder or uneven category use that might otherwise go unnoticed. Thus, the study contributes not only to SRL research but also to methodological advancements in the measurement of learner characteristics.

4.7.1. Limitations

The study involved only gifted students in accelerated programs in Indonesia, limiting generalizability to other student populations such as regular classrooms, international schools, or non-gifted learners. Self-regulated learning was measured using self-report scales, which are vulnerable to social desirability bias, overestimation, and subjective interpretation of response categories. Although item fit statistics were acceptable, several items displayed non-ideal thresholds. This suggests that the response scale may require refinement or the wording of items may need to be adjusted to improve category clarity. The study did not examine changes in SRL abilities over time. A longitudinal design could provide deeper insights into the developmental aspects of SRL among gifted students. The study focused on internal construct validity; external validity testing such as correlations with academic achievement, executive functioning, or teacher ratings was not conducted.

4.7.2. Implications for Behavioral Science

The validated SRL instrument offers a meaningful contribution to the existing body of knowledge by deonstrating that the self-regulation constructs among gifted students in accelerated programs follow a multidimensional framework consisting of pre-learning, during-learning, post-learning, and motivational components. These findings reinforce and extend current SRL theories such as Zimmerman’s and Pintrich’s models by showing that gifted learners may exhibit different patterns of cognitive and motivational regulation compared to general student populations. Moreover, the identification of threshold irregularities contributes to psychometric theory by highlighting developmental differences in how younger gifted learners interpret response categories. This insight supports the refinement of SRL measurement models, especially for populations with advanced cognitive abilities but varying emotional or metacognitive maturity.

Future research should expand the sample to include broader educational contexts, such as regular-track classes, international schools, or gifted programs outside Indonesia, to enhance external validity. Longitudinal studies are recommended to examine how SRL develops over time and to identify factors that strengthen or hinder self-regulation in accelerated settings. Additionally, mixed-method approaches incorporating interviews, observations, or teacher assessments can triangulate self-report data, reduce bias and improve instrument robustness. Further psychometric refinement is also necessary, particularly through response category revision or item rewording to resolve threshold disorders. Researchers may also explore relationships between SRL profiles and academic achievement, mental health, executive functioning, or learning motivation among gifted learners.

For educational practice, the validated instrument can serve as a diagnostic tool for teachers and counselors in madrasahs to identify strengths and weaknesses in students’ self-regulation skills. This information can guide the design of individual learning plans, mentoring strategies, and enrichment activities tailored to gifted learners’ needs. At the programmatic level, accelerated programs can use SRL assessments to design interventions that promote metacognitive awareness, learning autonomy, and motivation management skills essential for sustaining high academic performance.

From a policy perspective, the findings support the development of national guidelines for gifted education, particularly in Islamic educational environments such as madrasahs. Policymakers can integrate SRL assessment as part of student selection, monitoring, and evaluation systems in accelerated programs. Additionally, training policies for teachers could incorporate SRL-based pedagogical development to ensure educators are equipped to foster self-regulatory competence in high-ability learners. Ultimately, the instrument can contribute to data-driven policy decisions aimed at enhancing the quality, effectiveness, and accountability of gifted education programs at regional and national levels.

5. Conclusion

This study utilizes the Rasch Model to analyze the Self-Regulated Learning (SRL) scale in gifted students, with the aim of improving the accuracy and validity of the measurement. The Rasch model, assisted by R, was used to assess item suitability, item difficulty, and measurement structure consistency in junior high school students participating in an accelerated program at a madrasah. The results of the analysis show that this scale is able to accurately measure the SRL construct and has adequate psychometric qualities.

The SRL scale consists of two main subscales, namely independent learning skills and motivation. Independent learning skills include three phases: pre-learning (planning, time management, and learning environment management), learning process (information search, self-monitoring, and practice), and post-learning (self-evaluation and consequences for learning achievements). The motivation subscale covers aspects of task value, self-confidence, academic anxiety, attribution of failure, and goal orientation.

The results of the study show that Rasch analysis indicates that all items are within the ideal difficulty range (2 to 2), in accordance with the recommended criteria. However, a number of items, such as B1, B3, B5, B19, B36, and B67, were identified as requiring revision based on category threshold patterns that were not yet fully optimal.

These findings confirm the need to refine several items without reducing the overall suitability of the model. After linguistic and cultural adaptation, as well as adjustments to the characteristics of high-ability students in acceleration programs in Indonesia, 67 items were deemed suitable for use based on Rasch analysis.

This instrument is designed for madrasah students who are gifted and can complete the study in approximately 20–25 minutes, either online or offline. As a self-report-based measurement tool, this scale is effective for use in large-scale research; however, in this study, the researchers recommend that its use be supplemented with additional assessment techniques such as interviews or observations to produce a more comprehensive understanding. Next, the validity of this instrument is planned to be strengthened through advanced analysis approaches, such as Multiple Indicator Multiple Causes (MIMIC), Differential Item Functioning (DIF), Bayesian Approximation Measurement Invariance (BAMI), Measurement Invariance Explorer (MIE), or the application of Response Shift theory to identify changes in the self-regulation structure in students. Overall, the instrument analyzed through *difficulty*, *item fit*, and *Wright Map* proved to have good psychometric quality. The instrument is suitable for measuring the Self-Regulated Learning of outstanding students in SKS services, with minor improvements to items that are not yet fully in line with the model.

References

Aiyegbusi, O. L., Cruz Rivera, S., Roydhouse, J., Kamudoni, P., Alder, Y., Anderson, N., ... Calvert, M. J. (2024). Recommendations to address respondent burden associated with patient-reported outcome assessment. *Nature Medicine, 30*(3), 650–659.

Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing, 38*(1), 6–40.

Astuti, W., Hanafi, I., & Sarkadi. (2022). Program akselerasi belajar anak cerdas istimewa bakat istimewa. *Jurnal Ilmiah Pendidikan dan Pembelajaran (JIPP), 6*(3), 524–535. https://doi.org/10.23887/jipp.v6i3.54411

Banihashem, S. K., Bond, M., Bergdahl, N., Khosravi, H., & Noroozi, O. (2025). A systematic mapping review at the intersection of artificial intelligence and self-regulated learning. *International Journal of Educational Technology in Higher Education, 22*(1), Article 50.

Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge. https://doi.org/10.4324/9780429030499

Butler, D. L. (2023). Qualitative approaches to investigating self-regulated learning: Contributions and challenges. *Using Qualitative Methods to Enrich Understandings of Self-Regulated Learning: A Special Issue of Educational Psychologist, 1520*(January), 59–63. https://doi.org/10.4324/9781410608529-7

Carter, R. A., Rice, M., Yang, S., & Jackson, H. A. (2020). Self-regulated learning in online learning environments: Strategies for remote learning. *Information and Learning Sciences, 121*(5–6), 311–319. https://doi.org/10.1108/ILS-04-2020-0114

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management, 41*(2), 745–783.

Erdogan, T., & Senemoglu, N. (2016). Development and validation of a scale on self-regulation in learning. *SpringerPlus, 5*(1). https://doi.org/10.1186/s40064-016-3367-y

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., ... Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530.

Haerudin, & Nur, I. R. D. (2020). Analysis of student difficulties in inferential statistics courses. *Lebesgue Journal: Scientific Journal of Mathematics Education, Mathematics and Statistics, 1*(3), 208–215. https://doi.org/10.46306/lb.v1i3.35

Hertel, S., A, K. R., & B, Y. K. (2024). Machine translated by Google learning and instruction: Do independent learning profiles and intelligence mindsets relate to students' independent learning and achievement? *Machine Translated by Google, 90*(April 2023).

Larasati, P. E., Supahar, & Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. *Journal of Physics: Conference Series, 1440*(1). https://doi.org/10.1088/1742-6596/1440/1/012020

Marougkas, A., Troussas, C., Krouska, A., & Sgouropoulou, C. (2024). How personalized and effective is immersive virtual reality in education? A systematic literature review for the last decade. *Multimedia Tools and Applications, 83*(6), 18185–18233.

Melati, M., Mardiah, R., Safitri, N. T., & Thabran, Y. (2021). Self-directed learning sebagai alternatif pendekatan belajar di masa pandemi COVID-19 di SMPN 1 Merangin. *Jurnal Karya Abdi Masyarakat, 5*(3), 341–348. https://doi.org/10.22437/jkam.v5i3.16209

Mumpuni, K. E., Begimbetova, G. A., & Retnawati, H. (2023). Self-regulated learning questionnaire: Differential item functioning analysis using the Rasch model. *Varidika: Jurnal Penelitian dan Inovasi Pendidikan.*

Nabavi, R. T., & Bijandi, M. S. (2024). A literature review on Bandura's social learning theory and social cognitive learning theory.

Nisak, N., & Wibawa, S. C. (2021). Studi literatur efektivitas penerapan self-directed learning pada mata pelajaran pemrograman dasar berbasis Moodle. *IT-Edu: Jurnal Information Technology and Education, 5*(1), 283–293. https://doi.org/10.26740/it-edu.v5i1.37486

Nizaruddin, N., Muhtarom, M., Murtianto, Y. H., & Sutrisno, S. (2024). Examining the self-regulated learning scale using the Rasch model approach. *Indonesian Journal of Science and Mathematics Education, 5*(1), 518–530. https://doi.org/10.24042/ijsme.v5i1.21831

Qin, J., Qin, J., Qiu, J., Liu, Q., Li, M., & Ma, Q. (2024). SRL-ORCA: A socially aware multi-agent mapless navigation algorithm in complex dynamic scenes. *IEEE Robotics and Automation Letters, 9*(1), 143–150. https://doi.org/10.1109/LRA.2023.3331621

Ramadhani, R., Syahputra, E., & Simamora, E. (2024). The construct validity of self-regulated learning questionnaire for senior high school students: A Rasch model analysis. *International Journal of Evaluation and Research in Education, 13*(3), 2032.

Ridgley, L. M., DaVia Rubenstein, L., & Callan, G. L. (2022). Are gifted students adapting their self-regulated learning processes when experiencing challenging tasks? *Gifted Child Quarterly, 66*(1), 3–22. https://doi.org/10.1177/00169862211025452

Sari, A. D., Marniati, Rijanto, T., & Cholik, M. (2024). Model pembelajaran self-directed learning pada pelajaran pembuatan pola busana kerja siswa SMK ditinjau dari kemandirian belajar siswa. *Jurnal Ilmiah Pendidikan Citra Bakti, 11*(3), 917–928. https://doi.org/10.38048/jipcb.v11i3.3739

Sharma, K., Nguyen, A., & Hong, Y. (2024). Self-regulation and shared regulation in collaborative learning in adaptive digital learning environments: A systematic review of empirical studies. *British Journal of Educational Technology, 55*(4), 1398–1436.

Siekanska, M., Wilson, S. G., Blecharz, J., & Young, B. W. (2023). Examining the validity of the Polish short form version of the self-regulated learning-sport practice survey among competitive athletes. *Frontiers in Psychology, 14*, 1–10. https://doi.org/10.3389/fpsyg.2023.1132608

Singh, S. K., Del Giudice, M., Chiappetta Jabbour, C. J., Latan, H., & Sohal, A. S. (2022). Stakeholder pressure, green innovation, and performance in small and medium-sized enterprises: The role of green dynamic capabilities. *Business Strategy and the Environment, 31*(1), 500–514.

Subekti, A. S. (2021). Indonesian learners' self-directed learning and resilience in online English classes: Assessing interaction with L2 achievement. *Indonesian Journal of English Education, 8*(1), 1–16. https://doi.org/10.15408/ijee.v8i1.20681

Toering, T., Elferink-Gemser, M. T., Jonker, L., van Heuvelen, M. J. G., & Visscher, C. (2012). Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale. *International Journal of Sport and Exercise Psychology, 10*(1), 24–38. https://doi.org/10.1080/1612197X.2012.645132

Xu, Z., Zhao, Y., Liew, J., Zhou, X., & Kogut, A. (2023). Synthesizing research evidence on self-regulated learning and academic achievement in online and blended learning environments: A scoping review. *Educational Research Review, 39*, Article 100510.

Zimmerman, B. J. (2023). Dimensions of academic self-regulation: A conceptual framework for education. In *Self-regulation of learning and performance* (pp. 3–21). Routledge.