INDONESIAN JOURNAL OF SCIENCE AND MATHEMATICS EDUCATION
07 (3) (2024) 518-530
https://ejournal.radenintan.ac.id/index.php/IJSME/index
ISSN: 2615-8639
DOI: 10.24042/ijsme.v5i1.21831
November 2024

Examining the self-regulated learning scale using the Rasch model approach

Nizaruddin Nizaruddin[1], Muhtarom Muhtarom[2]*, Yanuar Hery Murtianto[3], Sutrisno Sutrisno[4]

[1,2,3,4]Mathematics Education Department, Universitas PGRI Semarang, Semarang, Indonesia

**Corresponding author:* muhtarom@upgris.ac.id

Article history:

Submitted: April 4, 2024
Accepted: November 5, 2024
Published: November 30, 2024

Keywords:

item response theory, Rasch model, scale, self-regulated learning

ABSTRACT

Self-regulated learning is a crucial aspect of the learning process for students. This ability is often overlooked due to the challenges of inaccurate measurement. This study aims to evaluate the quality of a self-regulated learning scale developed through an analysis of respondent responses. The research employed a descriptive quantitative approach using the Rasch Model as the analytical method. The instrument used consisted of 30 statement items. The study sample included 59 mathematics education students selected through cluster random sampling from two universities in different districts. The analysis results indicated that, after three calibration processes, the self-regulated learning scale was refined to 28 items with excellent quality. Furthermore, the responses of 58 students demonstrated a high level of consistency. Thus, self-regulated learning scale has good validity and reliability, making it a dependable tool for measuring self-regulated learning abilities. The implications of this study include the provision of a practical and reliable instrument for researchers and educators to support further studies and serve as an evaluation tool in learning development.

Mengkaji skala self-regulated learning menggunakan pendekatan model Rasch

Kata Kunci:

teori respons butir, model Rasch, skala, self-regulated learning

ABSTRAK

Self-regulated learning *merupakan aspek yang sangat penting dalam proses belajar siswa. Kemampuan ini sering terabaikan akibat kendala dalam pengukuran yang akurat. Penelitian ini bertujuan untuk menguji kualitas skala* self-regulated learning *yang telah dikembangkan melalui analisis kualitas respons responden. Penelitian menggunakan pendekatan kuantitatif deskriptif dengan Model Rasch sebagai metode analisis. Instrumen yang digunakan terdiri atas 30 item pernyataan. Sampel penelitian melibatkan 59 mahasiswa program studi pendidikan matematika yang dipilih secara* cluster random sampling *dari dua perguruan tinggi di kabupaten yang berbeda. Hasil analisis menunjukkan bahwa setelah tiga kali proses kalibrasi, skala* self-regulated learning *berhasil disempurnakan menjadi 28 item dengan kualitas yang sangat baik. Selain itu, respons 58 mahasiswa menunjukkan tingkat konsistensi yang tinggi. Dengan demikian, skala* self-regulated learning *memiliki validitas dan reliabilitas yang baik, sehingga dapat diandalkan sebagai alat ukur kemampuan* self-regulated learning. *Implikasi penelitian ini meliputi penyediaan instrumen yang praktis dan andal bagi peneliti maupun pendidik untuk mendukung studi lanjutan serta sebagai alat evaluasi dalam pengembangan pembelajaran.*

Contribution to the literature

This research contributes to:

- Strengthening the use of the Rasch model in validating educational instruments.
- Providing a measurement tool that can be used to evaluate self-regulated learning in various educational contexts.
- Offering a practical example of how the Rasch model can be applied to improve the quality of measurement instruments.

1. INTRODUCTION

Self-regulated learning is a concept of how a person becomes a manager of himself in his learning activities. Self-regulated learning is a person's ability to activate and encourage thinking (cognition), feelings (affection), and actions (actions) that have been planned systematically and repeatedly to achieve a goal in learning [1], [2]. Self-regulated learning involves four aspects, namely cognitive, affective, motivational, and behavioral, that lead to the individual's ability to adjust his actions and goals to achieve the desired results about changing environmental conditions [2]. Self-regulated learning is based on Bandura's assumption of the triadic theory of reciprocity. According to this theory, behavior occurs because there are three interrelated determinants, namely self, behavior, and environment [3]. Self-regulation in learning consists of several phases, namely the planning phase, where students perform task analysis, set goals, and plan behavior. Then, there is the performance or implementation phase, where students monitor and control behavior, and the last is the evaluation phase, where students will self-reflect based on feedback obtained [4]–[7].

Zimmerman [7] states that being a self-regulating learner means that learners are proactive in their efforts to learn because these learners can recognize their strengths and weaknesses and can determine task-related learning goals and strategies. The ability to self-regulate in this learning process requires students to be able to always monitor their behavior regarding the achievement of goals and then reflect on their behavior to determine the effectiveness of the learning they have done and strive to be better in the next learning [8]–[10]. The ability to self-regulate in the learning process plays an important role in education. In various studies, it has been found that self-regulation is significantly positively correlated with academic achievement [1], [6], [10]–[12]. Self-regulation in learning and its relationship to academic achievement It was found that self-regulation in learning has a significant positive relationship with academic achievement [13], [14]. Self-regulation in learning affects learning outcomes by helping learners acquire and retain knowledge in a structured and methodological way [15], [16].

Self-regulated learning is so important, but the development of adequate measuring tools has not followed it. Therefore, it is necessary to develop a quality self-regulated learning scale. The self-regulated learning measurement instrument was developed by Parantika and Astawan [17]. The approach used is Content Validity Ratio (CVR) analysis, and empirical validity is seen using the product moment correlation formula. Another research was conducted by Lee [18]. Data analysis to obtain a valid and reliable instrument uses a Confirmatory Factor Analysis (CFA) approach. Even though the studies above concluded that the self-regulated learning assessment instruments were valid and reliable in mathematics learning, CFA only tests the extent to which indicators in one measured construct represent the self-regulated learning construct, and cannot explain the quality of the research respondents.

The quality of the instrument can be related to the latest psychometric theory, which can facilitate the development of this instrument [19]–[21]. The theory is the Item Response Theory (IRT). There are several models in IRT, one of which is the One Parameter Logistics (1PL) model, with the parameter being the item difficulty level (bi). The most popular 1PL model used is the Rasch model [19], [21], [22]. The Rasch model appeared popularized by Georg Rasch, a mathematician from Denmark. Rasch found that the error of one test correlated with the error of another test. If this is compared, it is found that the opportunity to answer the questions correctly is the same when the students' abilities are compared to the level of difficulty of the questions [22].

From these findings, Rasch concluded that a person who has a higher ability will have a greater probability of answering the question correctly. The same is true for items. Items that have a higher level of difficulty have a lower probability of completing the item than the other items [20], [23]. If in the classical theory model, the observed score (x) is expressed in terms and e, then in the Rasch xi model, it functions as a function of the location of the respondent (θ) and the location of the item (δ). In the analysis of achievement tests, the location of the respondent is usually referred to as the respondent's ability level, and the location of the item is referred to as the item's difficulty level. One of the features of the Rasch model is that it does not depend on the sample used. The Rasch measurement simultaneously structured questions from the hardest to the easiest and the respondents from the highest to the lowest ability. Therefore, any inconsistency in the answers of the respondents (misfit) or unusual patterns (outliers) will be detected [22], [24]–[26].

Classical test theory relies heavily on samples and instrument items. In classical theory, the measurement process is focused on the apparent score (x). In the Rasch model, the data used is the probability score (P), which is the comparison between the correct answers and the number of questions given. The odds score is then converted into an odds value. Then, by entering the logarithm function, we can find the logit value using the following formula: [19], [22], [27]. This value is called the logit W-score or measure value. The logit value is scalable and can be used for various analyses. Another advantage of the Rasch model over other methods, especially from classical test theory, is the ability to predict missing data, which is based on a systematic response pattern head. In other models, we usually estimate the missing data with a value of zero (0).

In contrast, the Rasch model will produce a prediction that is the best possible value of the missing data. Thus, the data obtained seems to be complete and more accurate in subsequent statistical analysis. Therefore, this research aims to develop a self-regulated learning scale and test its quality using Rasch modeling.

Previous studies have explored the measurement of self-regulated learning using various approaches, including specific self-regulated learning interventions for elementary school students [28], self-regulated learning training programs aimed at improving academic performance, strategies, and motivation among college students [29], supporting students' self-regulated learning in online learning using artificial intelligence applications [30] and self-regulated learning in online learning environments [31]. However, most of these studies primarily focused on validity and reliability without thoroughly examining the quality of individual responses or detecting item and respondent misfits in detail. This study aims to develop a self-regulated learning scale that is not only valid and reliable but also capable of analyzing item quality and individual responses using the Rasch model. The findings of this study provide a significant contribution by offering a more accurate and reliable measurement tool while paving the way for further research involving larger populations to enhance the generalizability of the findings.

2. METHOD

The research employed a descriptive quantitative approach using the Rasch Model as the analytical method. Rasch modeling analysis is one of the psychometric techniques that is applied to improve the accuracy of the instrument construct, monitor the quality of the instrument, and calculate the respondent's ability [27]. Different from the classical theory, which focuses on group scales, the Rasch model considers every item of the scale, even down to the characteristics of the respondents working on it [22]. The Rasch model reveals the relationship between one's ability and item difficulty [23]. The raw data from the rating scale is converted into an "equal-interval scale" which is measured in logs (log odd units).

An item is said to be valid when it can distinguish between able respondents and those who are unable. There are two possibilities regarding this. The first possibility is the discrepancy of the respondents involved in the given exam. Rasch modeling can detect respondents who do not fit to be involved in data collection and can be excluded because they do not fit the existing model [26]. The second possibility is if it turns out that the item cannot distinguish the ability of respondents between those who are able and those who are unable. The item needs to be revised or even discarded. This clearly shows that Rasch's modeling not only measures the reliability of items but also tests the validity of the concept of the instrument used [22].

Reliability describes how steady the measurement results are. In the Rasch model, reliability is described by the existence of a separation index [26]. Separation reliability in the Rasch model reports two things, namely item reliability and person reliability. Separation reliability describes how far the measuring instrument can produce a measuring range on the logit ruler. After receiving ethics approval, the research was conducted. The separation reliability (item or person reliability) will be high if the research sample and item difficulty level have a wide range and produce small measurement errors. A broad item means that the item has a difficulty level from the easiest to the most difficult. Likewise, for the research sample, a wide sample means that the sample has abilities that are spread from the most intelligent to the least intelligent. Usually, low reliability is due to too few samples, so the hierarchical variation on the logit ruler is only slightly [21], [22]. The item separation index is an estimate of the distribution of grains on the measured variable. Reliability is said to be high if it produces a price above 3.00. According to Sumintono and Widhiarso [21], summary statistics provide overall information about the quality of respondents as a whole, the quality of the instruments used, and the interaction between person and item. After receiving consent from students, the research was conducted.

2.1 Person Measures

The average value that is more than logit 0.0 indicates the tendency of respondents who answer to agree more on statements in various items [21]. This suggests that the respondents generally found the items to align with their perspectives or experiences, reflecting a positive response trend. Such a tendency also implies that the items are effectively designed to resonate with the target population, enhancing the reliability and validity of the measurement tool.

2.2 Cronbach's Alpha

Cronbach's alpha value (measuring reliability, namely the interaction between person and item as a whole). These values can be interpreted using the guidelines presented in Table 1.

Table 1. Cronbach's Alpha Value

Value	Description
< 0.5	Bad
0.5 – 0.6	Ugly
0.6 – 0.7	Enough
0.7 – 0.8	Good
> 0.8	Very good

2.3 Person Reliability and Item Reliability

The values of person reliability and item reliability can be interpreted using the guidelines presented in Table 2.

Table 2. Person Reliability and Item Reliability Value

Value	Description
< 0.67	Weak
0.67 – 0.8	Enough
0.81 – 0.90	Good
0.91 – 0.94	Very good
> 0.94	Special

2.4 INFIT MNSQ and OUTFIT MNSQ

Other data that can be used are INFIT MNSQ and OUTFIT MNSQ; for the person table, the ideal value is 1.00 (the closer to 1.00, the better). For INFIT ZSTD and OUTFIT ZSTD, the average value in the person table is 0.0 (the closer the value is to 0.0, the better the quality), likewise for the item table [22], [26], [27].

2.5 Separation Value

The separation value shows the grouping of people and items. The greater the value of separation, the better the quality of the instrument in terms of overall respondents and items [21]. Another equation used to examine groupings more closely is strata separation.

$$H = \frac{\left[\left(4 \times SEPARATION\right) + 1\right]}{3} \quad (1)$$

According to Sumintono and Widhiarso [21], item fit order and person fit order provide info in checking whether an item or person is fit. This is done for screening so that misfit items or persons can be removed and instruments with good quality and consistent responses from respondents can be obtained. The INFIT MNSQ value of each item can be used to check the fit and misfit items. The mean and standard deviation values are summed and then compared; the logit value greater than this value indicates a misfit item. Other criteria according to Sumintono and Widhiarso [22] used to check the suitability of items that do not fit or misfits are: (1) Accepted Outfit Mean Square (MNSQ) value: 0.5 < MNSQ < 1.5; (2) Value of Outfit Z-Standard (ZSTD) accepted: -2.0 < ZSTD < +2.0; and (3) Point Measure Correlation (Pt Mean Corr) value received: 0.4 < Pt Measure Corr < 0.85. A fit item or person meets at least one of the above criteria.

3. RESULTS AND DISCUSSION

Sample selection was done by cluster random sampling. Respondents in the study were 59 students of mathematics education study programs at two universities from two different cities. In terms of gender, 16.95% of respondents were male and 83.05% female. In the first calibration using the Rasch Model, it was found that the quality of the instrument was very good, and the answers from the respondents were consistent. After a more in-depth look at each instrument item, it was found that two items were identified as misfits, namely item 4 and item 30. In addition, after a more in-depth look at each response from the respondents, there were 4 misfit persons, namely K09, K06, K22, and P07. For the misfit person, it is ignored first because it prioritizes the elimination of the misfit item. If all items are not misfit, then the misfit person can be eliminated [20], [21], [23], [24]. Therefore, it is necessary to calibrate the two Rasch Models on the research data after removing the two items.

In the second calibration using the Rasch Model, the quality of the instrument is very good, and the respondents' answers are consistent. After a more in-depth look at each instrument item, I found that all items were not indicated as misfits. However, after an in-depth look at each respondent's answer, one respondent, K06, was identified as a misfit. Therefore, it is necessary to calibrate the three Rasch Models on the research data after removing one respondent.

In the third calibration using the Rasch Model, it is known that the quality of the instrument is very good, and the answers from the respondents are very consistent. After a more in-depth look at each instrument item, I found that all items were not indicated as misfits. The same thing also happened after an in-depth look at each respondent's answer; no respondent was found to be identified as a misfit. Therefore, the research data is declared feasible for analysis because it is obtained from instruments that are of very good quality, and the results of the respondents' answers are very consistent [21]. The data of this research, which is the result of an ordinal scale Likert scale, is converted by the Rasch Model into a logit with a ratio scale, thus fulfilling the classical assumption of parametric statistics, namely the dependent variable on an interval or ratio scale. In detail, the explanation of the results of the Rasch Model analysis on this calibration is presented in Figure 1.

Figure 1 provides overall information about the quality of the respondents, the quality of the instruments, and the interactions between persons and items. Person measure = 0.96 indicates the tendency of respondents to answer frequently (on favorable items) and never (on unfavorable items) on statements on various items. Cronbach's alpha value measures reliability, namely the interaction between the person and the item as a whole. Cronbach's alpha value is classified as very good [21], [22]. The value of person reliability and the value of item reliability is classified as very good, so it can be concluded that the answers from the respondents are very consistent, and the quality of the items in the instrument is very good [21], [24], [26]. INFIT MNSQ and OUTFIT MNSQ for the table person, the average values are 1.02 and 1.00, respectively. The ideal value is 1.00 (the closer to 1.00, the better). For INFIT ZSTD and OUTFIT ZSTD, the average values are -0.05 and -0.09, respectively. The ideal value is 0.00 (the closer to 0.00, the better). Likewise, for the item table, the average values of INFIT MNSQ and OUTFIT MNSQ are 1.01 and 1.00, respectively. The ideal value is 1.00 (the closer to 1.00, the better). For INFIT ZSTD and OUTFIT ZSTD, the average value is -0.11 and -0.07, respectively. The ideal value is 0.00 (the closer to 0.00, the better). The grouping of people and items can be seen from the separation value [18]. The greater the value of separation, the better the quality of the instrument in terms of overall respondents and items because it can identify groups of respondents and groups of items. The value of person separation is 3.12, then H = 4.49 rounded up to 4, which means that there are four groups of respondents. The value of item separation is 4.11, then H = 5.81 rounded up to 6, which means that there are six groups of items.

These results highlight the high precision of the instrument in distinguishing between groups of respondents and items, confirming its effectiveness in capturing variability within the dataset. A person separation value of 3.12 and an item separation value of 4.11 indicate that the instrument can adequately differentiate among four distinct respondent groups and six item difficulty levels. This level of granularity is crucial for identifying patterns in respondent behavior and the relative challenge posed by different items, providing a deeper insight into the alignment of the instrument with the measured construct.

Additionally, the reliability indicators support the robustness of the instrument. The high person and item reliability values demonstrate consistency in responses and stability of item difficulty across different respondents. The near-ideal INFIT and OUTFIT statistics further confirm that the items function well within the expected parameters, ensuring that the measurement reflects the true abilities of respondents without significant bias. These findings validate the quality of the instrument and its readiness for application in broader educational contexts to assess the intended constructs accurately.

```
INPUT: 58 Person  28 Item  REPORTED: 58 Person  28 Item  4 CATS WINSTEPS 5.2.0.0
--------------------------------------------------------------------------------
      SUMMARY OF 58 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT          OUTFIT    |
|          SCORE      COUNT     MEASURE   S.E.      MNSQ   ZSTD    MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      79.5       28.0         .96    .32      1.02   -.05    1.00   -.09 |
|  SEM       1.5         .0         .15    .01       .06    .21     .05    .18 |
| P.SD      11.0         .0        1.13    .04       .43   1.55     .38   1.36 |
| S.SD      11.1         .0        1.14    .04       .43   1.56     .38   1.38 |
| MAX.     109.0       28.0        4.92    .62      2.91   5.05    2.34   3.31 |
| MIN.      39.0       28.0       -3.27    .30       .34  -3.43     .40  -3.02 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .35 TRUE SD    1.08  SEPARATION  3.12  Person RELIABILITY  .91 |
|MODEL RMSE    .32 TRUE SD    1.09  SEPARATION  3.40  Person RELIABILITY  .92 |
| S.E. OF Person MEAN = .15                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .99 (approximate due to missing data)
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .91
SEM = 3.27 (approximate due to missing data)
STANDARDIZED (50 ITEM) RELIABILITY = .95

      SUMMARY OF 28 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT          OUTFIT    |
|          SCORE      COUNT     MEASURE   S.E.      MNSQ   ZSTD    MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     164.6       58.0         .00    .22      1.01   -.11    1.00   -.07 |
|  SEM       3.9         .0         .19    .00       .06    .34     .05    .30 |
| P.SD      20.5         .0         .99    .02       .30   1.77     .27   1.56 |
| S.SD      20.9         .0        1.01    .02       .30   1.80     .28   1.59 |
| MAX.     218.0       58.0        1.55    .32      1.55   2.68    1.52   2.55 |
| MIN.     131.0       58.0       -3.01    .21       .47  -3.79     .50  -3.38 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .23 TRUE SD     .96  SEPARATION  4.11  Item   RELIABILITY  .94 |
|MODEL RMSE    .22 TRUE SD     .97  SEPARATION  4.41  Item   RELIABILITY  .95 |
| S.E. OF Item MEAN = .19                                                     |
-------------------------------------------------------------------------------
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00 (approximate due to missing data)
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

Figure 1. Summary Statistics

INPUT: 58 Person 28 Item REPORTED: 58 Person 28 Item 4 CATS WINSTEPS 5.2.0.0

Person: REAL SEP.: 3.12 REL.: .91 ... Item: REAL SEP.: 4.11 REL.: .94

Person STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
8	39	28	-3.27	.39	2.91	5.05	2.34	3.31	A .49	.46	71.4	67.8	K09
7	90	28	1.93	.32	2.01	3.29	1.86	2.70	B .59	.48	42.9	56.9	K08
46	90	28	1.93	.32	1.87	2.92	1.85	2.69	C .41	.48	39.3	56.9	P17
37	86	28	1.54	.31	1.72	2.48	1.65	2.21	D .58	.49	50.0	55.9	P08
19	92	28	2.14	.32	1.60	2.15	1.51	1.71	E .47	.47	50.0	57.7	K20
21	85	28	1.45	.31	1.41	1.57	1.54	1.91	F .42	.50	46.4	56.3	K22
28	61	28	-.78	.31	1.48	1.70	1.48	1.71	G .65	.52	50.0	58.1	K29
27	88	28	1.73	.31	1.44	1.68	1.39	1.44	H .30	.49	50.0	56.4	K28
56	73	28	.34	.30	1.41	1.49	1.39	1.43	I .51	.52	32.1	58.2	P27
12	74	28	.43	.30	1.40	1.48	1.38	1.40	J .58	.52	42.9	58.3	K13
49	86	28	1.54	.31	1.40	1.51	1.35	1.32	K .60	.49	32.1	55.9	P20
36	90	28	1.93	.32	1.38	1.46	1.33	1.24	L .43	.48	53.6	56.9	P07
24	75	28	.52	.30	1.28	1.08	1.25	.99	M .56	.52	57.1	58.4	K25
53	82	28	1.17	.30	1.22	.91	1.23	.92	N .42	.51	57.1	57.2	P24
2	109	28	4.92	.62	1.05	.27	1.21	.51	O .12	.24	89.3	89.5	K02
14	89	28	1.83	.31	1.17	.72	1.11	.49	P .59	.48	53.6	56.9	K15
5	72	28	.25	.30	1.09	.44	1.12	.52	Q .46	.52	50.0	57.9	K05
6	73	28	.34	.30	1.12	.52	1.12	.53	R .52	.52	60.7	58.2	K07
10	71	28	.15	.30	1.12	.52	1.10	.46	S .54	.52	64.3	57.5	K11
34	93	28	2.24	.33	1.11	.52	1.09	.41	T .38	.46	46.4	58.1	P05
30	77	28	.70	.30	1.08	.38	1.06	.31	U .28	.52	64.3	58.2	P01
44	83	28	1.26	.31	1.07	.36	1.04	.25	V .65	.50	46.4	56.8	P15
13	78	28	.80	.30	.97	-.02	1.02	.15	W .30	.51	57.1	58.0	K14
45	85	28	1.45	.31	1.02	.18	1.00	.08	X .43	.50	60.7	56.3	P16
51	72	28	.25	.30	.98	-.01	.99	.03	Y .43	.52	60.7	57.9	P22
29	70	28	.06	.30	.98	.00	.96	-.08	Z .71	.52	67.9	57.2	K30
BETTER FITTING NOT SHOWN													
16	75	28	.52	.30	.91	-.27	.89	-.38	z .63	.52	67.9	58.4	K17
48	67	28	-.22	.31	.91	-.25	.90	-.31	y .57	.52	57.1	56.8	P19
41	91	28	2.03	.32	.87	-.49	.84	-.54	x .68	.47	57.1	57.6	P12
47	71	28	.15	.30	.75	-.95	.87	-.44	w .26	.52	71.4	57.5	P18
52	83	28	1.26	.31	.87	-.45	.86	-.51	v .62	.50	53.6	56.8	P23
11	93	28	2.24	.33	.85	-.54	.85	-.50	u .47	.46	67.9	58.1	K12
31	72	28	.25	.30	.82	-.68	.85	-.52	t .19	.52	57.1	57.9	P02
57	94	28	2.35	.33	.84	-.58	.81	-.63	s .60	.46	53.6	59.3	P28
26	73	28	.34	.30	.81	-.70	.81	-.69	r .64	.52	60.7	58.2	K27
40	70	28	.06	.30	.81	-.70	.80	-.73	q .47	.52	60.7	57.2	P11
23	81	28	1.07	.30	.79	-.80	.79	-.80	p .65	.51	71.4	57.6	K24
35	60	28	-.88	.31	.74	-1.04	.75	-.99	o .35	.52	64.3	58.3	P06
18	69	28	-.03	.30	.73	-1.07	.74	-1.01	n .30	.52	60.7	57.2	K19
9	75	28	.52	.30	.73	-1.08	.73	-1.07	m .67	.52	71.4	58.4	K10
20	76	28	.61	.30	.73	-1.08	.73	-1.08	l .57	.52	64.3	58.3	K21
17	83	28	1.26	.31	.54	-1.46	.66	-1.43	k .13	.50	75.0	56.8	K18
54	73	28	.34	.30	.65	-1.45	.66	-1.41	j .59	.52	67.9	58.2	P25
32	86	28	1.54	.31	.65	-1.57	.64	-1.55	i .46	.49	67.9	55.9	P03
1	82	28	1.17	.30	.64	-1.58	.63	-1.57	h .46	.51	78.6	57.2	K01
4	72	28	.25	.30	.59	-1.78	.60	-1.70	g .54	.52	64.3	57.9	K04
42	86	28	1.54	.31	.58	-1.96	.57	-1.93	f .70	.49	67.9	55.9	P13
25	88	28	1.73	.31	.55	-2.11	.55	-2.01	e .63	.49	71.4	56.4	K26
39	74	28	.43	.30	.52	-2.21	.51	-2.24	d .68	.52	75.0	58.3	P10
50	70	28	.06	.30	.47	-2.50	.46	-2.56	c .66	.52	67.9	57.2	P21
38	81	28	1.07	.30	.45	-2.67	.45	-2.69	b .66	.51	85.7	57.6	P09
33	77	28	.70	.30	.34	-3.43	.40	-3.02	a .58	.52	75.0	58.2	P04
MEAN	79.5	28.0	.96	.32	1.02	-.05	1.00	-.09			60.0	58.4	
P.SD	11.0	.0	1.13	.04	.43	1.55	.38	1.36			11.8	4.6	

Figure 2. Person Statistics: Misfit Order

To check fit and misfit persons, the INFIT MNSQ value in Figure 2 of each person can be used; the mean and standard deviation values are added up and then compared; the logit value greater than this value indicates a misfit person. MEAN + SD = 1.43, so from this criterion, there are five people with an INFIT MNSQ value greater than 1.43, namely

K09, K08, P17, P08, and K20. Other criteria are used to check for non-conforming persons (outliers or misfits). The Outfit Mean Square (MNSQ) value received is 0.5 < MNSQ < 1.5. The acceptable Z-Standard Outfit (ZSTD) value is -2.0 < ZSTD < 2.0. The value of Point Measure Correlation (Pt Mean Corr) received is 0.4 < Pt Measure Corr < 0.85. Considering the other criteria, it can be concluded that there is no indication of a misfit person [21], [22].

INPUT: 58 Person 28 Item REPORTED: 58 Person 28 Item 4 CATS WINSTEPS 5.2.0.0

Person: REAL SEP.: 3.12 REL.: .91 ... Item: REAL SEP.: 4.11 REL.: .94

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
20	151	58	.64	.21	1.55	2.68	1.52	2.55	A .65	.55	48.3	57.6	i21
23	218	58	-3.01	.32	1.42	1.64	1.12	.43	B .23	.43	75.9	79.1	i24
21	172	58	-.30	.21	1.39	2.07	1.39	1.97	C .41	.54	48.3	56.4	i22
16	137	58	1.27	.21	1.36	1.81	1.37	1.84	D .24	.55	51.7	59.2	i17
14	169	58	-.16	.21	1.20	1.13	1.32	1.68	E .43	.54	55.2	56.1	i15
9	180	58	-.67	.22	1.22	1.24	1.27	1.38	F .29	.53	56.9	57.5	i10
11	138	58	1.22	.21	1.24	1.30	1.22	1.14	G .56	.55	50.0	59.2	i12
25	187	58	-1.01	.22	1.21	1.19	1.18	.93	H .52	.52	53.4	58.7	i26
3	164	58	.06	.21	1.19	1.08	1.18	1.00	I .44	.55	62.1	55.9	i3
28	167	58	-.08	.21	1.18	1.03	1.18	1.02	J .52	.54	53.4	56.0	i29
5	134	58	1.41	.22	1.15	.83	1.14	.76	K .42	.55	55.2	59.4	i6
8	149	58	.73	.21	1.15	.87	1.13	.73	L .51	.55	56.9	58.0	i9
12	179	58	-.62	.22	1.14	.85	1.12	.68	M .50	.53	56.9	57.5	i13
22	173	58	-.35	.21	1.12	.72	1.09	.52	N .64	.54	48.3	56.7	i23
18	173	58	-.35	.21	1.09	.59	1.11	.63	n .48	.54	55.2	56.7	i19
15	154	58	.50	.21	1.09	.54	1.07	.42	m .62	.55	60.3	57.2	i16
24	185	58	-.91	.22	1.05	.34	.99	.00	l .65	.52	60.3	58.3	i25
17	134	58	1.41	.22	.89	-.54	.85	-.75	k .64	.55	63.8	59.4	i18
4	159	58	.28	.21	.73	-1.64	.84	-.89	j .41	.55	56.9	56.4	i5
1	176	58	-.48	.21	.47	-3.79	.80	-1.13	i .52	.54	72.4	57.2	i1
10	194	58	-1.36	.23	.78	-1.28	.77	-1.13	h .59	.51	70.7	60.4	i11
27	173	58	-.35	.21	.78	-1.36	.76	-1.39	g .73	.54	56.9	56.7	i28
7	166	58	-.03	.21	.76	-1.47	.75	-1.49	f .73	.55	50.0	55.9	i8
6	187	58	-1.01	.22	.72	-1.75	.72	-1.56	e .68	.52	69.0	58.7	i7
13	151	58	.64	.21	.62	-2.37	.61	-2.43	d .71	.55	70.7	57.6	i14
2	131	58	1.55	.22	.57	-2.70	.56	-2.76	c .53	.55	70.7	59.7	i2
19	142	58	1.04	.21	.57	-2.72	.56	-2.79	b .66	.55	75.9	58.7	i20
26	167	58	-.08	.21	.50	-3.49	.50	-3.38	a .73	.54	75.9	56.0	i27
MEAN	164.6	58.0	.00	.22	1.01	-.11	1.00	-.07			60.0	58.4	
P.SD	20.5	.0	.99	.02	.30	1.77	.27	1.56			8.9	4.2	

Figure 3. Item Statistics: Misfit Order

To check fit and misfit items, Figure 3 can be used below, specifically the INFIT MNSQ value of each item; the mean and standard deviation values are summed and then compared; the logit value greater than this value indicates the item is a misfit. MEAN + SD = 1.31, so from this criterion, there are four items with a higher MNSQ INFIT value of 1.31, namely i21, i24, i22, and i17. Other criteria used to check for non-conforming items (outliers or misfits): The Outfit Mean Square (MNSQ) value received is 0.5 < MNSQ < 1.5. The acceptable Z-Standard Outfit (ZSTD) value is -2.0 < ZSTD < 2.0. The value of Point Measure Correlation (Pt Mean Corr) received is 0.4 < Pt Measure Corr < 0.85 [21], [22]. Considering the other criteria, it can be concluded that there is no indication of a misfit item.

Scale validation is important before the assessment because the instrument used must be valid first. If not, then the credibility and accuracy of the measurement will not be strong [23]–[25], [27]. Therefore, this research focuses on analysis. Through Rasch modeling, the scale validation carried out becomes more detailed because it reveals not only the items but also the participants. The analysis of Rasch modeling in this study focused on the fit to Rasch Model test, item analysis, and person analysis. One of the strengths of this study is that it uses Rasch modeling to uncover measurements that are not easy to perform using traditional analytical methods [21]-[32]. In addition, the research sample was obtained using a cluster random sampling technique from the population in Central Java Province.

This study aims to address the need for valid and reliable measurement of self-regulated learning, as identified in previous research. Wolters and Brady [33] revealed that although self-regulated learning significantly influences students' academic success, existing measurement instruments often fail to encompass all critical dimensions of SRL, such as planning, monitoring, and evaluation. Furthermore, Crede and Phillips [34], in their meta-analysis, noted that most SRL measurement instruments rely solely on traditional approaches, disregarding individual response quality or inconsistencies in the data. In this context, this study offers a solution by utilizing the Rasch model to develop an SRL scale capable of thoroughly evaluating item and response quality. This scale not only strengthens the validity and reliability of the instrument but also contributes new insights to the development of measurement tools that can be broadly applied in higher education contexts. The results of this study are expected to address the need for instruments capable of providing in-depth and comprehensive analysis of students' SRL abilities.

This study highlights the advantages of using the Rasch model, which enables in-depth analysis of the validity and reliability of the scale, including the identification of misfit items and respondents. The findings indicate that the scale possesses high validity and reliability, making it suitable for further research. However, the study has limitations, such as a limited sample size and a lack of demographic diversity among respondents, which may restrict the generalizability of the findings. Additionally, the removal of misfit items and respondents may reduce data diversity. To overcome these limitations, future research is recommended to involve larger and more diverse samples, as well as conduct a more detailed analysis of misfit patterns to understand the factors influencing responses. Developing instruments that can be applied across disciplines or cultures, along with integrating other approaches such as qualitative analysis, could also provide more comprehensive insights into self-regulated learning.

4. CONCLUSION

The results of this study indicate that a very good self-regulated learning scale with 28 items was obtained through three calibrations. This is also supported by the very consistent quality of the responses from 58 people. This self-regulated learning scale is of good quality and feasible to use. Other researchers who want to research self-regulated learning can use this scale and try it out with a much larger number of respondents. The implications of this research include providing practical instruments for researchers and educators to further study.

AUTHOR CONTRIBUTION STATEMENT

NN contributed to the fieldwork, the preparation of the background, and the supervision of the entire article's writing. MM contributed to the fieldwork and the data entry. YHM contributed to the fieldwork, and data documentation. SS contributed to the data analysis, the interpretation, and the literature review, and provided critical contributions to the manuscript's drafting and revision.

REFERENCES

[1] B. J. Zimmerman, "A social cognitive view of self-regulated academic learning," *J. Educ. Psychol.*, vol. 81, no. 3, pp. 329-339, 1989, doi: 10.1037//0022-0663.81.3.329

[2] B. J. Zimmerman, "Self-regulated learning and academic achievement: An overview," *Educ. Psychol.*, vol. 25, no. 1, pp. 3–17, 1990, doi: 10.1207/s15326985ep2501_2

[3] E. A. Locke, "Self-efficacy: The exercise of control," *Pers. Psychol.*, vol. 50, no. 3, pp. 801-804, 1997.

[4] S. Abbasnasab, M. R. Mohd Saad, and R. Boroomand, "Self-Regulated Learning Strategies (SRLS) and academic achievement in pre-university EFL learners," *Calif. Linguist. Notes*, vol. 37, no. 1, pp. 1–35, 2012.

[5] B. Bai, J. Wang, and Y. Nie, "Self-efficacy, task values and growth mindset: What has the most predictive power for primary school students' self-regulated learning in english writing and writing competence in an asian confucian cultural context?," *Cambridge J. Educ.*, vol. 51, no. 1, pp. 65–84, 2021, doi: 10.1080/0305764X.2020.1778639

[6] D. H. Schunk and B. J. Zimmerman, "Self-regulation in education: Retrospect and prospect," in *Self-regulation of learning and performance*, Routledge, 2023, pp. 305–314, doi: 10.4324/9780203763353-13

[7] B. J. Zimmerman, "Attaining self-regulation: A social cognitive perspective," in *Self-regulation: Theory, research, and applications/Academic*, 2000, ch. 2, pp. 13–39. doi: 10.1016/B978-012109890-2/50031-7

[8] D. L. Butler, "Qualitative approaches to investigating self-regulated learning: Contributions and challenges," in *Using Qualitative Methods To Enrich Understandings of Self-regulated Learning*, Routledge, 2023, pp. 59–63, doi : 10.4324/9781410608529-7

[9] B. J. Zimmerman, "Dimensions of academic self-regulation: A conceptual framework for education," in *Self-regulation of learning and performance*, Routledge, 2023, pp. 3–21, doi: 10.4324/9780203763353-1

[10] B. J. Zimmerman, D. Greenberg, and C. E. Weinstein, "Self-regulating academic study time: A strategy approach," in *Self-regulation of learning and performance*, Routledge, 2023, pp. 181–199, doi ; 10.4324/9780203763353-8

[11] N. L. Adam, F. B. Alzahri, S. C. Soh, N. A. Bakar, and N. A. M. Kamal, "Self-regulated learning and online learning: A systematic review," in *Advances in Visual Informatics: 5th International Visual Informatics Conference, IVIC 2017, Bangi, Malaysia, November 28–30, 2017, Proceedings 5*, 2017, pp. 143–154, doi : 10.1007/978-3-319-70010-6_14

[12] M. Raković *et al.*, "Examining the critical role of evaluation and adaptation in self-regulated learning," *Contemp. Educ. Psychol.*, vol. 68, no. 1, pp. 1-14, 2022, doi: 10.1016/j.cedpsych.2021.102027

[13] 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

[14] C. Schuster, F. Stebner, D. Leutner, and J. Wirth, "Transfer of metacognitive skills in self-regulated learning: An experimental training study," *Metacognition Learn.*, vol. 15, no. 3, pp. 455–477, 2020, doi: 10.1007/s11409-020-09237-5

[15] A. Pandey, D. Hale, S. Das, A.-L. Goddings, S.-J. Blakemore, and R. M. Viner, "Effectiveness of universal self-regulation–based interventions in children and adolescents: A systematic review and meta-analysis," *JAMA Pediatr.*, vol. 172, no. 6, pp. 566–575, 2018, doi : 10.1001/jamapediatrics.2018.0232

[16] D. A. Robson, M. S. Allen, and S. J. Howard, "Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review," *Psychol. Bull.*, vol. 146, no. 4, pp. 1-91, 2020, doi: 10.1037/bul0000227

[17] I. W. A. Parantika and I. G. Astawan, "Pengembangan instrumen self regulated learning dan kecerdasan emosional pada pembelajaran matematika kelas V sekolah dasar," *PENDASI J. Pendidik. Dasar Indones.*, vol. 6, no. 2, pp. 133–140, 2022, doi: 10.23887/jurnal_pendas.v6i2.1357

[18] M. Lee, "Development the measurement instrument of online self-regulated learning in higher education," *J. Educ. Technol.*, vol. 34, no. 4, pp. 901–927, 2018, doi: 10.17232/KSET.34.4.901

[19] S. W. Chan, Z. Ismail, and B. Sumintono, "A Rasch model analysis on secondary students' statistical reasoning ability in descriptive statistics," *Procedia-Social Behav. Sci.*, vol. 129, no. 12, pp. 133–139, 2014, doi: 10.1016/j.sbspro.2014.03.658

[20] J. Fan and T. Bond, "Applying Rasch measurement in language assessment: Unidimensionality and local independence," in *Quantitative Data Analysis for Language Assessment Volume I*, Routledge, 2019, pp. 83–102, doi : 10.4324/9781315187815-5

[21] B. Sumintono and W. Widhiarso, *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*. Jawa Barat: Trim Komunikata Publishing House, 2014.

[22] B. Sumintono and W. Widhiarso, *Aplikasi pemodelan Rasch pada assessment pendidikan*. Jawa Barat: Trim komunikata, 2015.

[23] T. G. Bond and C. M. Fox, *Applying the Rasch model: Fundamental measurement in the human sciences*. London: Psychology Press, 2013.

[24] O. Huei, R. Rus, and A. Kamis, "Construct validity and reliability in content knowledge of design and technology subject: A Rasch measurement model approaches for pilot study," *Int J Acad Res Bus Soc Sci*, vol. 10, no. 3, pp. 497–511, 2020, doi : 10.6007/IJARBSS/v10-i3/7066

[25] E. B. Tunç, "Assessing the psychometric properties of the brief resilience scale: A Rasch modeling approach," *Int. J. Eurasian Educ. Cult.*, vol. 8, no. 23, pp. 2116-2137, 2023.

[26] S. Yasin, M. F. M. Yunus, and I. Ismail, "The use of Rasch measurement model for validity and reliability," *J. Couns. Educ. Technol.*, vol. 1, no. 2, pp. 22–27, 2018.

[27] W. J. Boone, "Rasch analysis for instrument development: Why, when, and how?," *CBE—Life Sci. Educ.*, vol. 15, no. 4, pp. 1-7, 2016, doi: 10.1187/cbe.16-04-0148

[28] M. Lee, S. Y. Lee, J. E. Kim, and H. J. Lee, "Domain-specific self-regulated learning interventions for elementary school students," *Learn. Instr.*, vol. 88, pp. 1-18, 2023, doi: 10.1016/j.learninstruc.2023.101810

[29] M. Theobald, "Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis," *Contemp. Educ. Psychol.*, vol. 66, no. 1, pp. 1-84, 2021, doi: 10.1016/j.cedpsych.2021.101976

[30] S.-H. Jin, K. Im, M. Yoo, I. Roll, and K. Seo, "Supporting students' self-regulated learning in online learning using artificial intelligence applications," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, pp. 1-21, 2023, doi: 10.1186/s41239-023-00406-5

[31] R. A. Carter Jr, M. Rice, S. Yang, and H. A. Jackson, “Self-regulated learning in online learning environments: Strategies for remote learning,” *Inf. Learn. Sci.*, vol. 121, no. 5, pp. 321–329, 2020, doi: 10.1108/ILS-04-2020-0114

[32] L. Prieto, J. Alonso, and R. Lamarca, “Classical test theory versus rasch analysis for quality of life questionnaire reduction,” *Health Qual. Life Outcomes*, vol. 1, no. 1, pp. 1–13, 2003, doi: 10.1186/1477-7525-1-27

[33] C. A. Wolters and A. C. Brady, “College students’ time management: A self-regulated learning perspective,” *Educ. Psychol. Rev.*, vol. 33, no. 4, pp. 1319–1351, 2021, doi: 10.1007/s10648-020-09519-z, doi: 10.1007/s10648-020-09519-z

[34] M. Credé and L. A. Phillips, “A meta-analytic review of the motivated strategies for learning questionnaire,” *Learn. Individ. Differ.*, vol. 21, no. 4, pp. 337–346, 2011, doi: 10.1016/j.lindif.2011.03.002, doi: 10.1016/j.lindif.2011.03.002