The Rasch Analysis Approach for Validating The Precision of Mathematical Cognitive Load Scale

Rina Oktaviyanthi[1,*] and Ria Noviana Agus[1]

Department of Mathematics Education, Universitas Serang Raya, Serang-Banten, Indonesia
*Corresponding author. rinaokta@unsera.ac.id

Abstract. High cognitive load in learning mathematics occurs partly due to the mental activity of working memory experiences fullness and maximum pressure during the learning process. This situation intensified by the characteristics of the information, for instance, when studying the formal definition of limits involving high-level mathematical formulations and constructing epsilon and delta propositions. Previously, to determine the cognitive load on students, it has been traced by cognitive load scales. However, there hasn't been much improvement of cognitive load scales specific to the characteristics of abstract mathematical concepts as epsilon and delta proof. Hence, the aim of this research is to validate a specific cognitive load scale for the formal definition of limits concept and examine item instrument consistency. Rasch analysis applied as the research approach through aspects of reliability testing, separation, and item measurement. The Winstep software employed as a tool to explore the psychometric properties of the instrument. A total of 75 students from three private universities in Serang, Cilegon, and Pandeglang participated as research respondents. The main findings confirm that the tested cognitive load scale meets the valid Rasch model assumptions and is reliable with fairly good and consistent reliability and separation value categories.

Keywords: rasch analysis, mathematical cognitive load scale, limits concept

1 Introduction

The concept of function limits at the university level presents quite a high challenge, especially for first-year students [1][2]. The challenge of the concept of function limits in question is the formal definition of limits, which involves the abstract mathematical ideas of epsilon and delta [3][4]. This understanding of epsilon and delta extends beyond the context of limits and is integral to calculus as a whole. It is inseparable in proving theorems in Advanced Calculus and Real Analysis. The ability to use the concepts of epsilon and delta is regarded as a fundamental skill that students need to master in order to model changes in variable movements, validate mathematical arguments, identify theorem-proof requirements, or justify a proposition [5][6][7]. This further emphasizes the importance of learning the concepts of epsilon and delta as a support for understanding calculus and as a stimulus for advanced mathematical thinking.

In practical, mental fatigue is often experienced by students when studying abstract mathematical concepts, including epsilon and delta. This diagnosis is evident

H. S. Panigoro et al. (eds.), *Proceedings of the 2nd International Conference on Sciences, Mathematics, and Education 2023 (ICOSMED 2023)*, Advances in Social Science, Education and Humanities Research 927,
https://doi.org/10.2991/978-2-38476-410-5_46

from the observations in the preliminary study, which included the supervision of student response questionnaires during the course of learning. The results of the response questionnaires revealed that students frequently experience reduced concentration, decreased information absorption, difficulties in solving problems, and some students even complained of headaches or tension during their studies. These symptoms displayed by students occur when individuals experience high cognitive load [8][9][10]. Cognitive load can be described as an individual's cognitive capacity or the thinking capacity when processing received information within a specific timeframe [11][12]. This mental activity evolves within an individual's working memory, the stability of which needs to be maintained to remain normal. Increased cognitive load can impact students' performance in constructing knowledge and the understanding process [13][14][15]. Therefore, it is crucial to recognize situations where cognitive load can increase and to manage it effectively to maintain both performance and mental healthiness.

In principle, regarding information processing as stated in the cognitive load theory, it is identified there is a limitation on the amount of information possibly effectively processed by an individual's cognitive system at any given time. This theory suggests that activities related to knowledge construction and problem-solving require mental resources [12][16][17]. High cognitive load occurs when the amount of mental effort required exceeds the available resources, leading to difficulties in information processing and learning activities. Not only the characteristics of mathematical concepts (intrinsic cognitive load) contribute to the increase in students' cognitive load, but also the instructional design (extraneous cognitive load) in terms of how the material is presented and the level of mental effort (germane cognitive load) exerted by students can influence cognitive load [12][18][19]. Thus far, to determine the presence of cognitive load in students, cognitive load scales have been used. However, there have been limited developments in cognitive load scales specific to the characteristics of abstract mathematical concepts, such as the epsilon and delta proof. Therefore, the aim of this research is to validate a cognitive load scale specific to the concept of formal limit definition and to investigate the item consistency in the instrument.

The research growth in constructing and evaluating cognitive load scales has remained stable over the past five years. Among these studies, there is Liao (2019) investigates cognitive load in the utilization of instructional videos and collaborative digital gaming [20]. Then, Janssen's (2020) evaluates the perspective of Collaborative Cognitive Load Theory (CCLT) in computer-supported collaborative learning (CSCL) research [21]. There is also Andersen (2021) tests the cognitive load instrument in virtual learning environments through investigations of the validity of measurement using the Partial Credit Model (PCM), Confirmatory Factor Analysis (CFA), and retention test correlations [22]. Additionally, Benaim (2022) develops a cognitive assessment scale for stroke patients by verifying the psychometric properties of the scale using principal component analysis [23]. Furthermore, Nori (2023) measures cognitive load scales in spatial navigation as a discovery in updating the Environmental Knowledge Model (EKM) to explain human navigation concerning individual and environmental factors [24]. Several of these studies indicate the urgency of investigating cognitive load and the continued relevance of cognitive load scale development. However, the availability of cognitive load scales that focus on the complex epsilon-delta concept is very limited.

Based on the comparison between the ideal conditions and the real-world situation, as well as the urgency of the need for measuring cognitive load, especially in the abstract context for formal definition of limit concept, the purpose of this research is to validate a specific cognitive load scale for the formal definition of limit concept and to investigate the consistency of items in the instrument.

2 Method

A quantitative research approach using the Rasch model was implemented to validate and investigate the item consistency of the cognitive load scale. The Rasch model tests utilized included reliability, separation, and item measurement [25][26][27]. Data analysis was computed using WINSTEPS software version 5.6.2 to examine the psychometric aspects of the cognitive load scale. The criteria for reliability and separation tests are presented in Table 1 below.

Table 1. Reliability and Separation Criteria [28]

Criteria	Poor	Fair	Good	Very Good	Excellent
Person and item measurement reliability	< 67	67–80	81–90	91–94	> 94
Person and item strata separated	2 or less	2–3	3–4	4–5	> 5

The research data was collected through the completion of a cognitive load scale by the respondents. This cognitive load instrument is in the form of a 17-item rating scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree) and is divided into three aspects of cognitive load. The individual statements for each aspect are distinguished by the codes: ICL for statements related to the characteristics of the material concept, ECL for statements emphasizing the delivery design of the material, and GCL for statements focusing on the mental efforts of the students in responding to the material.

The study involved a total of 75 first-year university students from three private universities in Banten. The respondents were not selected with any specific criteria except that they were currently enrolled in a Calculus course. The respondents exhibited heterogeneity in terms of gender and mathematical ability levels. This approach was intended to ensure the objectivity of the data analysis results.

3 Result and discussion

The results of the reliability and separation analyses for the cognitive load scale related to the formal definition of limit in the WINSTEPS software can be found in the output for Summaries of Person and Items, as presented in Figure 1 below.

PERSON	75 INPUT	75	MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	46.6	17.0	.05	.37	.98	-.2	.99	-.1
P.SD	2.3	.0	.27	.04	.43	1.5	.44	1.4
REAL RMSE	.37	TRUE SD	.00	SEPARATION 5.03		PERSON RELIABILITY .94		

ITEM	17 INPUT	17	MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	205.8	75.0	.00	.17	.99	-.2	.99	-.1
P.SD	94.8	.0	2.36	.02	.24	1.8	.23	1.7
REAL RMSE	.17	TRUE SD	2.35	SEPARATION 7.47		ITEM RELIABILITY .99		

Fig. 1. WINSTEPS Output for Summaries of Person and Items

In Figure 1, the average person measure score (top blue box) is 0.05, while the average item measure score (bottom blue box) is 0.00. The person measure score is higher than the item measure score, indicating that the respondents' abilities are greater compared to the difficulty of the cognitive load scale items related to the formal definition of limit [29][30]. Furthermore, we can examine the values of person reliability indicates the consistency of respondents' answers, and item reliability reflects the quality of the statements in the instrument [31][32]. The higher the values of person and item reliability in a scale, the better the scale can measure the consistency of respondents' responses and the quality of the items on that scale. Figure 1 shows that the person and item reliability values for the cognitive load scale are 0.94 and 0.99, respectively. Based on Table 1, the person reliability is categorized as excellent, and the item reliability is considered excellent as well.

Another aspect that supports assess the adequacy of a scale is the INFIT and OUTFIT MNSQ and ZSTD values. The criterion for INFIT and OUTFIT MNSQ is the closer values to 1, the more ideal the measured scale is [33]. The criterion for the ideal value of a scale based on ZSTD is that approaches to 0 [34]. Referring to this criterion, Figure 1 reveals that the INFIT and OUTFIT MNSQ values for both the person and item aspects are around 0.98 and 0.99. Additionally, the ZSTD values for both the person and item aspects are around -2 and -1. The meaning of these MNSQ values approaching 1 and ZSTD values tending towards 0 implies that the cognitive load scale related to the formal definition of limit meets the ideal standards of an instrument.

Next is the aspect of separation for both person and item. Separation of responses focuses on how well a set of items in an instrument can cover a range of logit abilities. The higher the person separation value, the more ideal an instrument is because the statement items in it can span persons with various abilities, from low to high levels [35][36]. The person separation value in Figure 1 is 5.03, which falls into the good category. This means that the statements in the cognitive load scale related to the formal limit definition cover a wide range of ability levels (a continuum) from difficult to easy statements.

On the other hand, item separation emphasizes how widely a sample subjected to measurement can be distributed along a linear interval scale. The higher the item separation value, the more ideal the measurement is [37]. This index is also useful for describing the significance of the construct being measured [38]. Based on Figure 1, the item separation value is 7.47 categorized as excellent. This implies that the respondents' data meet the standard of ideal variation in measuring a scale, resulting in data from respondents with both high and low cognitive loads.

The results of the item validity analysis for the cognitive load scale connected to the formal definition of limit in the WINSTEPS software are displayed in output for Item Measures, as presented in Figure 2 below.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	ITEM
15	210	75	.32	.10	GCL5
4	212	75	.30	.10	ICL4
16	214	75	.28	.10	GCL6
3	226	75	.17	.10	ICL3
1	229	75	.14	.10	ICL1
2	236	75	.07	.10	ICL2
5	236	75	.07	.10	ICL5
11	239	75	.04	.10	GCL1
13	239	75	.04	.10	GCL3
12	240	75	.03	.10	GCL2
14	252	75	-.08	.10	GCL4
9	253	75	-.09	.10	ECL4
7	255	75	-.11	.10	ECL2
17	264	75	-.20	.10	GCL7
8	266	75	-.23	.10	ECL3
10	270	75	-.27	.10	ECL5
6	288	75	-.47	.11	ECL1
MEAN	242.9	75.0	.00	.10	
P.SD	21.1	.0	.21	.00	

Fig. 2. WINSTEPS Output for Item Measures

Based on the analysis in Figure 2, it is observed that the standard deviation value is 0.21. The difficulty levels of items in the cognitive load scale linked to the formal definition of limit can be categorized based on the standard deviation and the logit values in the MEASURE column. An item is considered very difficult if the logit value is > +1SD, difficult if the logit value falls in the range of 0.00 logit to +1SD, easy if the value is in the range of 0.00 logit to -1SD, and very easy if the logit value is <-1SD. According to this categorization, the following thresholds for each category were determined: 1) very difficult items have logit values > 0.21 (GCL5, ICL4, GCL6), 2) difficult items have logit values between 0.00 and 0.21 (ICL3, ICL1, ICL2, ICL5, GCL1, GCL3, GCL2), 3) easy items have logit values between 0.00 and (-0.21) (GCL4, ECL4, ECL2, GCL7), and 4) very easy items have logit values < -0.21 (ECL3, ECL5, ECL1). With the varying difficulty levels of items that can be grouped into different categories, it strengthens the indication that the developed instrument capable to accommodate respondents with different ability levels, which is one of the indicators of a good instrument [39][40].

4 Conclusion

Based on the processing and data analysis by the Rasch model approach, it has been confirmed that the cognitive load scale, focused on students' difficulties in the formal definition of limit concept, meets the valid Rasch model assumptions and is reliable. This is determined by aspects of reliability, separation, and instrument item consistency. The person and item reliability values fall into the ”very good” and ”excellent” categories with indices of 0.94 and 0.99. Meanwhile, the separation values for persons and items are categorized as ”good” and ”very good” with indices of 5.03 and 7.47. Furthermore, in terms of item difficulty levels, four categories have been established that can encompass varying levels of respondent abilities in the study. Thus, the cognitive load scale for formal limit definition can serve as a reference instrument to measure the extent of cognitive load experienced by students when studying this abstract concept.

References

1. A. Ayebo, A. Mrutu, International Electronic Journal of Mathematics Education **14**(2), 385–392, (2019)
2. T. Mkhatshwa, Mathematical Thinking and Learning **22**(2), 139–161, (2020)
3. S. Seager, PRIMUS **30**(1), 88–96, (2020)
4. G. Hill, T. More, Journal of Higher Education Theory & Practice **21**(5) (2021)
5. B. Faulkner, N. Johnson-Glauch, D. Choi, G. Herman, Journal of engineering Education **109**(3), 402–423, (2020)
6. T. Mingus, M. Koelling, PRIMUS **31**(3-5), 393–412, (2021)
7. T. Bedada, M. Machaba, EURASIA Journal of Mathematics, Science and Technology Education **18**(10) (2022)
8. V. Markova, T. Ganchev, K. Kalinkov, in *affect and stress recognition in 2019 International Conference on Biomedical Innovations and Applications (BIA* (IEEE, 2019-11), pp. 1–4 ,
9. K. Hochberg, S. Becker, M. Louis, P. Klein, J. Kuhn, Journal of Science Education and Technology **29**, 303–317, (2020)
10. M. Dubravac, B. Meier, Quarterly journal of experimental psychology **76**(7), 1515–1538, (2023)
11. J. L., S. Plass, Educational Psychology Review **31**, 339–359, (2019)
12. J. Sweller, Educational Technology Research and Development **68**(1), 1–16, (2020)
13. A. Lai, C. Chen, G. Lee, British Journal of Educational Technology **50**(1), 232–247, (2019)
14. C. Wang, T. Fang, Y. Gu, Computers & Education **143, 103683** (2020)
15. J. Buchner, K. Buntins, M. Kerres, Journal of Computer Assisted Learning **38**(1), 285–303, (2022)
16. D. Feldon, G. Callan, S. Juth, S. Jeong, Educational Psychology Review **31**, 319–337, (2019)
17. F. Paas, J. Merriënboer, Current Directions in Psychological Science **29**(4), 394–398, (2020)
18. S. Sepp, S. Howard, S. Tindall-Ford, S. Agostinho, F. Paas, Educational Psychology Review **31**, 293–317, (2019)
19. A. Skulmowski, K. Xu, Educational psychology review **1-26** (2021)
20. C. Liao, C. Chen, S. Shih, Computers & Education **133**, 43–55, (2019)
21. J. Janssen, P. Kirschner, Educational Technology Research and Development **68**(2), 783–805, (2020)
22. M. Andersen, G. Makransky, Journal of Computer Assisted Learning **37**(1), 183–196, (2021)

23. C. Benaim, G. Wauquiez, D. Pérennou, C. Piscicelli, B. Lucas-Pineau, H. Bonnin-Koang, H. Devilliers, Physical and Rehabilitation Medicine **65**(3) (2022)
24. R. Nori, M. Zucchelli, M. Palmiero, L. Piccardi, Journal of Environmental Psychology **88, 102032** (2023)
25. S. Chan, C. Looi, W. Ho, W. Huang, P. Seow, L. Wu, Heliyon **7**(9) (2021)
26. S. Adeniji, P. Baker, (Mathematics Education Research Group of Australasia, 2022)
27. W. Romine, A. Lannin, M. Kareem, N. Singer, in *Advances in Applications of Rasch Measurement in Science Education* (Springer International Publishing, Cham, 2023), pp. 427–453 ,
28. W. Fisher, Jr, Journal of Interdisciplinary Economics **35**(1), 29–60, (2023)
29. J. Park, E. Park, Frontiers in psychology **10, 1992** (2019)
30. P. Jefferies, M. Ungar, PloS one **15**(9) (2020)
31. A. Adeniran, Scholar Journal of Applied Sciences and Research **2**(4), 1–5, (2019)
32. L. Sürücü, A. Maslakci, Business & Management Studies: An International Journal **8**(3), 2694–2726, (2020)
33. M. Canto-Cerdan, P. Cacho-Martinez, F. Lara-Lacarcel, A. García-Munoz, Scientific Reports **11**(1) (2021)
34. A. Alhadabi, S. Aldhafri, European Journal of Educational Research **10**(2), 957–973, (2021)
35. M. Planinic, W. Boone, A. Susac, L. Ivanjek, Physical Review Physics Education Research **15**(2) (2019)
36. M. Chan, R. Subramaniam, Rasch measurement: Applications in quantitative educational research **159-178** (2020)
37. K. Ahorsu, C. Lin, V. Imani, M. Saffari, M. Griffiths, A. Pakpour, International journal of mental health and addiction **1-9** (2020)
38. S. Saidi, N. Siew, International Electronic Journal of Mathematics Education **14**(3), 535–546, (2019)
39. L. Clark, D. Watson, Psychological assessment **31**(12) (2019)
40. M. Peeters, J. Augustine. Currents in pharmacy teaching and learning (2023)