Infinity Journal of Mathematics Education
Volume 13, No. 1, February 2024
p-ISSN 2089-6867
e–ISSN 2460-9285
https://doi.org/10.22460/infinity.v13i1.p99-118

COGNITIVE LOAD SCALE IN LEARNING FORMAL DEFINITION OF LIMIT: A RASCH MODEL APPROACH

Rina Oktaviyanthi[1*], Ria Noviana Agus[1], Mark Lester B. Garcia[2], Kornkanok Lertdechapat[3]
[1]Universitas Serang Raya, Indonesia
[2]Ateneo de Manila University, Philippines
[3]Chulalongkorn University, Thailand

Article Info

Article history:

Received Nov 29, 2023
Revised Jan 19, 2024
Accepted Jan 20, 2024
Published Online Jan 23, 2024

Keywords:

Cognitive load scale,
Formal definition of limit,
Item analysis,
Rasch model,
Unidimensionality

ABSTRACT

Constructing proofs for the limit using the formal definition induces a high cognitive load. Common assessment tools, like cognitive load scales, lack specificity for the concept of limits. This research aims to validate an instrument tailored to assess cognitive load in students focused on the formal definition of limits, addressing the need for diverse strategies in education. The research employs a quantitative survey design with a Rasch model approach, utilizing a data collection instrument in the form of a questionnaire. Subsequently, the data are analyzed by focusing on three aspects: (1) item fit to the Rasch model, (2) unidimensionality, and (3) rating scale. A total of 315 students from three private universities in Banten participated as research respondents. The findings of this study affirm the validity of the cognitive load scale centered on the formal definition of limit, meeting the stringent standards set by Rasch modeling. Additionally, the results of the study provide evidence of the scale's adherence to the monotonic principle of the Rasch model. These outcomes contribute to a comprehensive understanding of cognitive load in the context of learning formal definition of limit, providing a solid foundation for instructional design and assessment strategies.

Corresponding Author:

Rina Oktaviyanthi,
Department of Mathematics Education,
Universitas Serang Raya
Jl. Raya Cilegon Km. 5, Taktakan, Serang City, Banten 42162, Indonesia.
Email: rinaokta@unsera.ac.id

How to Cite:

Oktaviyanthi, R., Agus, R. N., Garcia, M. L. B., & Lertdechapat, K. (2024). Cognitive load scale in learning formal definition of limit: A rasch model approach. *Infinity*, *13*(1), 99-118.

1. INTRODUCTION

Formal definition of limit idea in the Calculus perspective plays a crucial role in the promoting students' higher-order thinking skills (Kidron, 2020; Parr, 2023; Thompson & Harel, 2021). The fundamental of abstractness in formal definition of limit takes a role initiative transmitter of advanced mathematical thinking to support students' critical reasoning and logical argumentation (Ghedamsi & Lecorre, 2021; Jablonka, 2020; Oktaviyanthi et al., 2018). Furthermore, the involvement of the epsilon-delta analytical context in the formal definition of limit aims to stimulate students' deductive thinking, optimizing their metacognitive potentiality and enabling them to solve problems reflectively (Brown, 2022; Case & Speer, 2021; Yan et al., 2020). Learning formal definition of limit concept not only fosters students' thinking skills concerning limits framework but also facilitates a rigorous and structured approach to honing their cognitive abilities as the foundation for higher-order mathematical thinking (Alam, 2020; Chen et al., 2021; Martínez-Planell & Trigueros, 2021; Viirman et al., 2022). Thus, instructing students to the abstract concept within formal definition of limit should be examined in learning to encourage and optimize mathematical thinking skills.

One significant challenge in achieving the ideal learning outcomes for the formal definition of limit places an emphasis on students constructing proofs processes engaging with epsilon and delta (Arzarello & Soldano, 2019; Quarfoot & Rabin, 2022; Slavíčková & Vargová, 2023). As an illustration, when students are asked to identify the value of delta for any positive epsilon given, they should be able to determine the required delta for any given epsilon. The proof typically begins with the statement "assume epsilon is greater than zero." However, to comprehend the proof, the procedure does not commence with epsilon, rather it initiates with delta (Adiredja, 2021; Oktaviyanthi et al., 2018). Students frequently find this type of proposition construction perplexing and exasperating. This situation leads to cognitive challenges for students, particularly as their working memory persistently processes information until they attain comprehension (Ludyga et al., 2022; Sepp et al., 2019). In the event that comprehension is unachieved within a specified timeframe, it may result in mental fatigue and the experience of maximum pressure. This leads to unmanageable over cognitive load. The excessive cognitive load experienced by students can have serious implications for the continuity of the learning process and the outcomes of teaching performance (Anmarkrud et al., 2019; Chew & Cerbin, 2021). Hence, it becomes priority imperative to closely monitor the students cognitive load dynamics throughout their learning process. Specifically for the concept of formal limit definition. Understanding it heavily relies on comprehending notations such as the inequality symbol, the absolute value, variables and Greek letters that act as constants all at the same time. Similarly, this requires student's prior understanding of elementary types of functions, the shape and behaviour of their graphs, as well as their defining characteristics. This comprises such as extensive set of prerequisite knowledge and skills that will place a heavy cognitive demand on students learning the formal definition of limit for the first time. Instructors need to identify the specific cognitive demands of the concept to understand how complex the concept is to learn and how it contributes to cognitive load. This understanding is crucial for instructors as a guide in designing materials and instructional methods so that students can not only manage their cognitive potential optimally but also have a focused and easily understandable learning experience.

Cognitive load can be assessed over various methods including brain activity measures (functional MRI), physiological measurements (electrical skin conductivity), eye tracking, self-rating scales, and even cardiovascular metrics (Can et al., 2019; Katona, 2022; Ramakrishnan et al., 2021; Skulmowski, 2023). The most commonly approach employed is the utilization of cognitive load scale instruments (Huckaby et al., 2022; Jiang & Kalyuga, 2020; Klepsch et al., 2017). Cognitive Load Theory considers three distinct categories of cognitive load namely cognitive load caused by the complexity of concepts (intrinsic cognitive load), cognitive load influenced by the design of learning materials (extraneous cognitive load), and cognitive load based on the cognitive activation processes of learners during learning (germane cognitive load) (Klepsch & Seufert, 2020; Skulmowski & Xu, 2022). In its development, cognitive load scales encompass all elements of cognitive load types to investigate which aspect of cognitive load is most dominantly perceived by students. However, the lack of detailed or widespread measurement tools related to students' cognitive load when facing the concept of limits leaves instructors without diverse strategies to address students' cognitive load. Therefore, it is necessary to develop a specified cognitive load scale for students' difficulties in understanding the formal definition of limit that meets the standards of instrument validity. The development of a cognitive load scale specifically for the formal limit concept is crucial not only to determine the extent of cognitive load experienced by students but also as an assessment measure of which cognitive aspects students feel burdened by and serves as a reference in meeting the learning needs of students with different knowledge backgrounds and cognitive capacities.

The improvement of cognitive load scales has been a relatively stable and enduring research subject over the past decade. Leppink et al. (2013) analyzed the primary components of cognitive load instruments to trace the effectiveness and efficiency of various learning environments based on learning strategies and student characteristics. Hadie and Yusoff (2016) validated the cognitive load scale in problem-based learning through construct validity and internal consistency tests using confirmatory factor analysis. Gupta and Zheng (2020) verified an instrument to assess students' cognitive load in mathematical problem-solving influenced by three factors: learning strategies, task difficulty, and prior knowledge. Ouwehand et al. (2021) confirmed the validity of four subjective cognitive load assessment scales: numeric scales (Likert and visual analog), and pictorial scales (facial emoticons and embodied) to measure which cognitive processes are more accurately represented. Huckaby et al. (2022) formulated a post-training cognitive load scale to measure knowledge acquisition and identify barriers in training. Krieglstein et al. (2023) developed and validated a cognitive load questionnaire for the three types of cognitive load through two analytical methods: principal component analysis and confirmatory factor analysis. From the review of previous research, apparently there is still a limited specificity in cognitive load scales related to the complexity for the formal definition of limit concept using the Rasch model approach.

Hence, in light of the immediacy and complexities associated with learning the formal definition of limits may lead to extreme cognitive load in students, the purpose of this research is to evaluate the validity of the student cognitive load instrument in comprehending the formal definition of limits to reduce the symptoms of disproportionate cognitive load that possibly will obstruct the achievement of teaching and learning objectives.

2. METHOD

The research objective is to assess the validity of a specialized student cognitive load instrument related to the formal definition of limits to mitigate the effects of excessive cognitive load. The research employs a quantitative survey design with a Rasch model approach, utilizing a data collection instrument in the form of a questionnaire (Faradillah & Febriani, 2021; Pradipta et al., 2021). The instrument being assessed is a Likert scale with five rating levels, comprising scale 1 (Strongly Disagree), scale 2 (Disagree), scale 3 (Neutral), scale 4 (Agree), and scale 5 (Strongly Agree). The instrument aspects investigated are divided into three types of Cognitive Load: Intrinsic Cognitive Load (ICL), Extraneous Cognitive Load (ECL), and Germane Cognitive Load (GCL) (Klepsch & Seufert, 2020; Krieglstein et al., 2023; Sweller, 2011) with statement items tailored to the research purpose, specifically in the context of the formal definition of limit as presented in Table 1.

Table 1. Cognitive load scale component

Cognitive Load Type	Cognitive Load Scale in Formal Definition Limit	Item
Intrinsic Cognitive Load (ICL)	The formal limit concept is challenging to comprehend	ICL1
	The explanation of formal limit is difficult to understand	ICL2
	The content of formal limit concept is highly complicated	ICL3
	The formal limit concept involves various complex information	ICL4
	Without prerequisite materials and prior knowledge, the formal limit concept remains incomprehensible	ICL5
Extraneous Cognitive Load (ECL)	Learning media aids in providing an overview the structure of formal limit concept	ECL6
	Learning media design in formal limit concept makes it challenging to recognize the relationships between concepts	ECL7
	Learning media design in formal limit concept facilitates thinking	ECL8
	Learning media design in formal limit concept makes it challenging to quickly locate relevant information	ECL9
	Learning media design assists in focusing the formal limit concept	ECL10
Germane Cognitive Load (GCL)	I actively visualize the formal limit concept	GCL11
	Learning media encourage me to actively think about the formal limit concept	GCL12
	I strive to understand the formal limit concept	GCL13
	I struggle to integrate each information section about the formal limit concept in the learning media to a more comprehensive and general concept	GCL14
	I find it challenging to fully understand the formal limit concept	GCL15
	I encounter difficulty in extending my prior knowledge for the formal limit concept under study	GCL16
	I am able to promptly and precisely apply the formal limit concept through learning media	GCL17

In Table 1, the components of statements in the cognitive load scale developed in this study are illustrated. There are a total of 17 items with 3 main item codes representing each type of cognitive load. The code ICL is for statements representing cognitive load related to the complexity of the formal limit definition concept, code ECL for items indicating cognitive load originating from the instructional design of how the formal limit definition concept is delivered, and code GCL for items identifying cognitive load based on the cognitive processes students engage in when understanding the formal limit definition concept and constructing comprehensive knowledge.

The statement items were validated by 12 experts with a minimum of 10 years of professional expertise, consisting of 3 experts in mathematics education, 2 mathematics experts, 3 experts in cognitive psychology, 2 experts in educational psychology, and 2 experts in the Indonesian language (Oktaviyanthi & Agus, 2023). The information on the instrument validation results using Fleiss Kappa Statistics (Falotico & Quatto, 2015; Gwet, 2021; Landis & Koch, 1977) is provided in Table 2.

Table 2. Validation results of cognitive load instrument in the formal limit context

Cognitive Load Type	Item	Fleiss K Index	Index Interpretation	Decision
Intrinsic Cognitive Load (ICL)	ICL1	0.47	Moderate agreement	Applicable
	ICL2	0.77	Substantial agreement	Applicable
	ICL3	0.45	Moderate agreement	Applicable
	ICL4	0.83	Almost perfect	Applicable
	ICL5	0.58	Moderate agreement	Applicable
Extraneous Cognitive Load (ECL)	ECL6	0.68	Substantial agreement	Applicable
	ECL7	0.81	Almost perfect	Applicable
	ECL8	0.83	Almost Perfect	Applicable
	ECL9	0.65	Substantial agreement	Applicable
	ECL10	0.62	Substantial agreement	Applicable
Germane Cognitive Load (GCL)	GCL11	0.84	Almost Perfect	Applicable
	GCL12	0.55	Moderate agreement	Applicable
	GCL13	0.70	Substantial agreement	Applicable
	GCL14	0.88	Almost perfect	Applicable
	GCL15	0.76	Substantial agreement	Applicable
	GCL16	0.49	Moderate agreement	Applicable
	GCL17	0.85	Almost perfect	Applicable

The cognitive load scale for understanding the formal limit definition was tested on 315 first-year students at three private universities in Banten, Indonesia who were enrolled in Calculus courses. Respondents for the study were not specifically selected to ensure the most objective results. Respondent characteristics were heterogeneous, encompassing both male and female students with varying levels of mathematical proficiency, including high, moderate, and low abilities. Information on the mathematical abilities of the students was obtained from a combination of the academic advisor's remarks, mathematics course grades, and confirmed through a pre-requisite test on the limit function chapter such as function concept and operations, equations and inequalities, quadratic and rational functions. The summary of the demographic distribution of the respondents is presented in Table 3.

Table 3. Demographic distribution of research respondents in the first semester

Region	Major Study	Gender		Total Subjcet	
Serang	Mathematics Education	M	10		
		F	135	M: 35	190
	Science and Engineering	M	25	F: 155	
		F	20		
Cilegon	Mathematics Education	M	5		
		F	40	M: 20	70
	Science and Engineering	M	15	F: 50	
		F	10		
Pandeglang	Mathematics Education	M	5		
		F	25	M: 15	55
	Science and Engineering	M	10	F: 40	
		F	15		
Total Subject				315	

A quantitative approach using the Rasch analysis model was applied to achieve the research objectives. WINSTEPS software version 5.6.2 was used as a data computational tool to support Rasch analysis. The instrument's validity test content using the Rasch model focuses on three aspects: (1) the adequacy of data with the tested mathematical model, in this case, the Rasch model (item fit analysis), (2) the instrument's item capabilities in measuring cognitive load (unidimensionality analysis), and (3) respondents' understanding of the instrument's rating scale differences (rating scale analysis) (Hadžibajramović et al., 2020; Toland et al., 2021; Yamashita, 2022).

Item fit analysis is the evaluation of the quality of statement items' suitability with a mathematical model and is detected as an indication of measurement validity in the Rasch model. Item fit analysis refers to how far the research data's distance is from its predicted values (Chan et al., 2021; Stenner et al., 2023). The technical item fit analysis is theoretically performed by examining the compatibility between the research data transformed into a specific function and the prediction function of the Rasch model. The evaluation of item fit suitability can use purely quantitative and visual approaches. The results of item fit analysis in the WINSTEPS software can be found in output table 10, with the following criteria (Silvia et al., 2021; Yamashita, 2022):

- Acceptance of Outfit Mean Square (MNSQ) values ranging from 0.5 < MNSQ < 1.5
- Acceptance of Outfit Z-standard (ZSTD) values ranging from -2.0 < ZSTD < +2.0
- Acceptance of Measure Correlation (PTMEASURE CORR) values ranging from 0.4 < PTMEASURE < 0.85

As for the graphical (visual) approach in WINSTEPS, it can be found in the GRAPSH menu with the Expected Score ICC option. The suitability of item fit in the graphical approach is assessed by how well the distribution of responses from the research data aligns with the Rasch model graph (Chi et al., 2023).

Unidimensionality analysis is the evaluation of the measurement quality of statement items in the research instrument, intended to measure only one attribute or construct.

Unidimensionality analysis is a validity aspect of the instrument that assesses how accurately it measures what it is intended to measure (Indihadi et al., 2022; Josa & Aguado, 2020). The results of unidimensionality analysis in WINSTEPS can be found in output table 23, categorized based on Fisher's criteria in Table 4 (Ramesh & Sanampudi, 2022).

Table 4. Fisher's unidimensionality criteria

Cognitive Load Type	Poor	Fair	Good	Very Good	Excellent
Variance in data explained by measure	< 50%	50 – 60%	60 – 70%	70 – 80%	> 80%
Unexplained variance in contrasts 1-5 of PCA of residuals	> 15%	10 – 15%	5 – 10%	3 – 5%	< 3%

The interpretation of unidimensionality relies on two criteria in Table 4. A higher variance value and a lower unexplained variance value indicate that the instrument is better at measuring the intended construct (Al Ali & Shehab, 2020). This means that the developed instrument does not have the potential to measure aspects beyond what it is intended to measure.

Rating scale analysis is the verification of ratings in the instrument to demonstrate the accuracy of the choices provided by respondents (Naar et al., 2021; Qu et al., 2023). Rating scale analysis serves to detect indications of respondents' misunderstanding of the rating of statement items. The results of rating scale analysis in WINSTEPS are shown in output table 3.2 in the columns of observed average (OBSVD AVRGE) and Andrich Threshold. The rating indicator of the instrument is considered to function properly if the values in the OBSVD AVRGE column follow the monotonic principle, increasing from rating 1 to rating 5, and the gap between ratings in the Andrich Threshold column ranges from 1.4 to 5.0 (Eckes & Jin, 2021; Swain et al., 2023).

3. RESULT AND DISCUSSION

3.1. Results

3.1.1. Item Fit Analysis

The item fit analysis results for the cognitive load scale focusing on formal limit definition in the WINSTEPS software are presented in output table 10, as depicted in Figure 1.

In Figure 1, it can be observed that the first criterion for instrument item validity is located in the Outfit Mean Square (MNSQ) column with an accepted range of values ranging from 0.5 < MNSQ < 1.5. The MNSQ values for the 17 statements in the cognitive load scale range from 0.65 to 1.47, indicating compliance with the MNSQ criterion. Furthermore, the second criterion for item validity is located in the Outfit Z-standard (ZSTD) column with an acceptance range of -2.0 < ZSTD < +2.0. The ZSTD values for each statement in the cognitive load scale fall within the interval of -1.98 to 1.83, indicating the achievement of the ZSTD criterion. As for the third criterion in item validity, it is examined in the Measure Correlation (PTMEACORR) column with an acceptance range of values from 0.4 < PTMEACORR < 0.85. The PTMEACORR values of the statements in the cognitive load scale fall within the parameter acceptance range, with the smallest value being 0.41 and the largest being 0.64.

INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
1.52	1.92	1.47	1.83	A .47	.18	21.3	43.2	GCL14
1.29	1.79	1.26	1.55	B .46	.18	37.3	45.9	ECL8
1.27	1.55	1.26	1.44	C .52	.17	48.0	51.6	GCL11
1.24	1.44	1.24	1.44	D .59	.17	44.0	50.0	ICL2
1.20	1.50	1.19	1.43	E .51	.19	32.0	43.8	GCL15
1.10	.80	1.10	.78	F .44	.19	38.7	44.5	ECL7
1.04	.27	1.05	.32	G .54	.15	68.0	65.7	ICL4
1.00	.07	1.00	.06	H .64	.21	38.7	38.0	GCL17
.99	-.03	.99	-.02	I .53	.18	49.3	51.4	ECL6
.92	-.61	.92	-.61	h .41	.19	49.3	44.2	ECL9
.88	-.96	.89	-.92	g .62	.21	46.7	38.1	ECL10
.86	-.76	.87	-.71	f .49	.16	69.3	60.9	ICL5
.80	-.98	.80	-.79	e .64	.21	50.7	38.0	GCL13
.73	-1.15	.72	-1.51	d .53	.20	53.3	38.0	GCL12
.69	-1.55	.69	-1.53	c .47	.18	56.0	43.2	ICL1
.69	-1.63	.69	-1.62	b .62	.20	53.3	44.2	GCL16
.64	-1.79	.65	-1.98	a .51	.19	60.0	42.2	ICL3
.99	-.24	.99	-.27			48.0	46.1	
.24	1.79	.23	1.73			11.9	7.6	

Figure 1. Output table 10 in WINSTEPS

In light of the completion of the MNSQ, ZSTD, and PTMEACORR parameters, it can be said that the statements in the cognitive load scale are in accordance with the predictive function of the Rasch model, and there is no conflict between the items and the measured construct (Chi et al., 2021; Koskey et al., 2017). Furthermore, PTMEACORR values > 0.40 indicate that the developed cognitive load scale has excellent item discrimination.

3.1.2. Unidimensionality Analysis

The results of unidimensionality analysis for the cognitive load scale of the formal limit definition in the WINSTEPS software are presented in output table 23, as shown in Figure 2.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	73.0841	100.0%		100.0%
Raw variance explained by measures =	56.0841	76.7%		76.3%
Raw variance explained by persons =	.6021	.8%		.8%
Raw Variance explained by items =	55.4821	75.9%		75.4%
Raw unexplained variance (total) =	17.0000	23.3%	100.0%	23.7%
Unexplned variance in 1st contrast =	4.1915	5.7%	24.7%	
Unexplned variance in 2nd contrast =	2.1981	3.0%	12.9%	
Unexplned variance in 3rd contrast =	1.5940	2.2%	9.4%	
Unexplned variance in 4th contrast =	1.4435	2.0%	8.5%	
Unexplned variance in 5th contrast =	1.2332	1.7%	7.3%	

Figure 2. Output table 23 in WINSTEPS

In Figure 2, it can be observed that the percentage of observed raw variance explained is 76.7%, which falls into the category of excellent according to the assumption of unidimensionality criteria. This value indicates that the developed instrument originates from a one-dimensional measurement or has excellent accuracy in measuring only one factor, which is the cognitive load of the formal limit definition. Additionally, the value of unexplained variance in the 1st contrast is 5.7%, which falls into the good category, reinforcing that the items in the instrument can differentiate the measured factor and are not influenced by other factors that should not be measured.

3.1.3. Rating Scale Analysis

The results of the rating scale analysis for the cognitive load scale of the formal limit definition in the WINSTEPS software are presented in output table 3.2, as shown in Figure 3.

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	339	27	-2.63	-2.62	.98	.98	NONE	(-3.64)	1
2	2	299	23	-2.30	-2.29	.88	.92	-2.34	-1.84	2
3	3	220	17	-.10	-.14	1.04	1.04	-1.27	.09	3
4	4	184	14	2.07	2.13	.79	.79	1.55	1.85	4
5	5	233	18	2.49	2.45	.94	.95	2.07	(3.47)	5

Figure 3. Output table 3.2 in WINSTEPS

In Figure 3, with the description "OBSVD AVRGE," it indicates the fulfillment of the monotonic principle, which means a consistent increase in values between ratings from 1 to 2 and so on. This increase serves as an indicator that the ratings in the cognitive load scale of the formal limit definition function effectively in eliciting responses from the research subjects. Another indication of whether the respondents understand each rating in the cognitive load scale is shown through a graphical approach, as depicted in Figure 4.

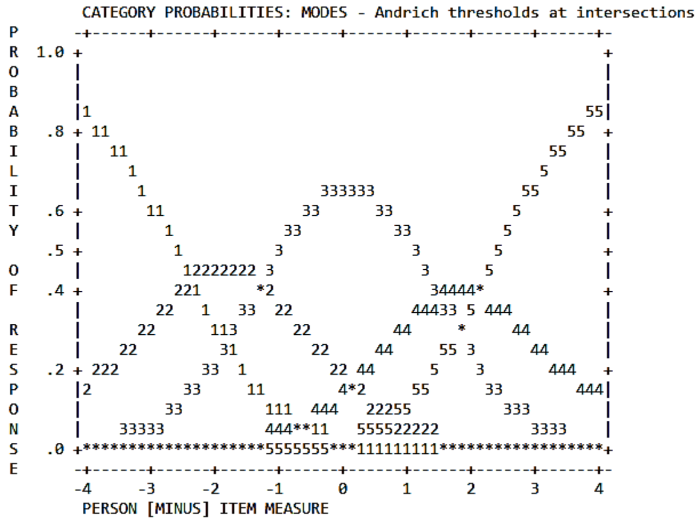

Figure 4. Rating test scale (partial-credit) graph

From the illustration in Figure 4, it can be observed that each rating has distinct peaks separated from one another. This fact indicates that the ratings in the cognitive load scale are understood by the respondents. Ratings 1, 3, and 5 have higher peaks than ratings 2 and 4. This information needs to be further examined through the calculation of the Andrich Threshold index.

The Andrich Threshold index is used to investigate how well and how accurately research subjects understand the options provided in the cognitive load scale. The ideal Andrich Threshold values range from NONE, negative, to positive, and the ideal distance between ratings falls within the logit values of 1.4 – 5.0. The detailed distances between ratings are shown in Table 5.

Table 5. Distance between Rating

Rating	Label	Andrich Threshold	Distance Between Rating	Distance Between Rating Type
1	SD	None		
			2.34	Ideal
2	D	-2.34		
			1.07	Non Ideal
3	N	-1.27		
			2.82	Ideal
4	A	1.55		
			0.52	Non Ideal
5	SA	2.07		

From Table 5, it can be observed that the non-ideal rating distances are present between rating 1 to rating 2 and rating 4 to rating 5. This non-ideality refers to a rating distance of < 1.4 logit. This fact indicates that respondents experienced confusion or did not accurately understand when selecting between rating 1 or 2 and 4 or 5. As pointed out in the works of Andrich (2013) and Andrich and Pedler (2019) the distance between non-ideal ratings illustrates respondents' misunderstanding in distinguishing rating choices according to the factual conditions they experience. Respondents assume that selecting an item with the 'Strongly Disagree' scale can be represented by ratings 1 or 2, but in reality, the difference between rating 1 and rating 2 has a different scale meaning, indicating a priority scale within their choices.

3.2. Discussion

The research findings emphasize the validity of the cognitive load scale based on the Rasch model to detect an increase in students' cognitive load when studying the formal limit concept. The first testing criterion is the analysis of item fit to determine the extent to which the instrument items provide appropriate information. In Rasch modeling, the validity of an item's statements is measured by parameters such as MNSQ, ZSTD, and PTMEACORR. Based on the analysis results presented in Figure 1, it can be concluded that the cognitive load scale items meet the validity criteria for parameters. Meeting these criteria means that the research subjects' response patterns to the instrument's statement items align with the prediction function of the Rasch model (Bond & Fox, 2007; De Ayala, 2018).

An interesting aspect of instrument validity testing using the Rasch model is the ability to trace the difficulty level of statements, from those considered the most difficult to those that are easiest for respondents to agree with. In Figure 1, under the ITEM column, the top position indicates the statement item that is most challenging to measure cognitive load responses, which is item coded as GCL14, and the bottom position provides information about the statement item that is the easiest to measure cognitive load responses, coded as ICL3. In the "Item Statistics" column, the metric commonly employed to measure the difficulty level of an item is the "Difficulty" or "Location" parameter (θ) (Chung & Cai, 2021; Shi et al., 2023). If an item has a positive θ value, it indicates a challenging item, as a higher level of ability is required to answer it correctly. Conversely, if an item has a negative θ value, it signifies an easy item, as it can be answered proficiently by respondents with lower levels of ability (Mutiawani et al., 2022; Nima et al., 2020). Therefore, the topmost position in the "Item Statistics" column signifies the most challenging item to measure, as respondents with lower ability levels have a lower likelihood of answering it correctly. This insight provides an understanding of how well the instrument can differentiate among individuals with varying levels of ability, and this information can be valuable in optimizing instrument design or enhancing measurement quality (Aghekyan, 2020; Brzezińska, 2020).

Item with code GCL14, designated as the most challenging item to gain respondent agreement, falls under the code of germane cognitive load, which are the cognitive load resulting from cognitive processes relevant to the understanding of the material being studied and the knowledge construction process. If an individual does not have germane cognitive load, it means their working memory cannot organize, construct, elaborate, or integrate the concepts they are currently learning into long-term memory (Forsberg et al., 2021; Skulmowski, 2023). Conversely, individuals with issues related to germane cognitive load tend to attempt to connect new knowledge with their existing knowledge, hoping to store it in their memory (Klepsch & Seufert, 2020). The more an individual faces complexity in processing new knowledge, the greater the germane cognitive load they experience (Bishara, 2022; Szulewski et al., 2021). Upon further examination in the ITEM column of Figure 1, it is evident that of the five statement items that respondents found most challenging to agree with, three of them are coded as GCL. This raises the assumption that the research subjects do not tend to have germane cognitive load. This indicates a lack of cognitive activation in students during the learning process. This information provides a recommendation for researchers to further investigate the factors causing this phenomenon.

Furthermore, item with code ICL3, regarded as the item that respondents found easiest to agree with, falls under the category of intrinsic cognitive load, which is the cognitive load resulting from the complexity of the material. The concept of a formal limit definition is considered complex due to its involvement of deep mathematical understanding, particularly concerning epsilon-delta numbers (Adiredja, 2021; Oktaviyanthi et al., 2018). Epsilon is seen as a number used to express the precision level or the acceptable margin of error in measuring the distance between the actual value of a function and the expected value (Brown, 2022). The smaller the epsilon value, the tighter the allowable error limit. On the other hand, delta indicates how close the independent variable must get to a specific value for the limit result to meet the epsilon-specified error limit. Recognizing that research subjects easily agreed that the formal limit definition concept is highly complex signifies the potential occurrence of intrinsic cognitive load (Case & Speer, 2021). This information can be a consideration for instructors to present content material in a way that is most understandable for students.

Although there are statement items in the instrument that are the most challenging and easiest in measuring cognitive load responses, overall, there are no items that deviate or fail to meet the criteria. This is further supported by the graphical approach of the Expected Score ICC, as shown in Figure 5.

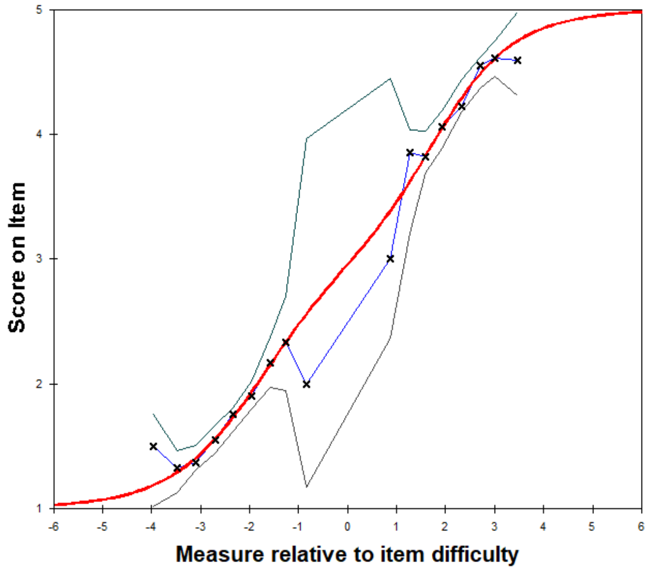

Figure 5. Expected score ICC graph

In Figure 5. the distribution of responses is represented by black crosses and connected by blue lines, spread in the confidence space of the Rasch prediction model, symbolized by the red line, and around the tolerance region of the curve, symbolized by the black line (Hoijtink, 2005; Wu et al., 2016). This visualization signifies that the 17 statements in the cognitive load scale of the formal limit definition can accurately measure what is intended to be measured. In more detail, Tesio et al. (2023) elaborate that the criteria for assessing a scale are as accurate as possible in depicting the latent variable being measured. This is observed from the expected score ICC graph, where the distribution of responses for each item aligns along the curve of the model, along with its tolerance boundaries. The ICC graph itself provides a representation of the expected scores by the Rasch model for each item on the y-axis at every level of the measurement continuum (x-axis) (Boone & Staver, 2020; Hagell, 2019).

Another research finding is related to the results of the Andrich Threshold index in Table 5, indicating an indication of respondents' inability to distinguish the options provided in the cognitive load scale. Non-ideal spacing between ratings occurs in options 2 (Disagree) and 4 (Agree). The ideal rating criteria are met when the rating scale analysis > 1.4 logit. If the logit value of the rating < 1.4, it is recommended to merge the rating scale. If the ratings are not merged, it signifies that the 5-Likert scale on the cognitive load instrument lacks adequate discrimination and has overlapping thresholds in various items, so the response format is considered for modification (Boone & Staver, 2020; Casale et al., 2023). Figure 6, part (a), shows the graph before merging the ratings, and part (b) displays the graph after merging the ratings. After merging rating 2 with rating 1 and rating 4 with rating 5, the logit values of the ratings become ideal according to the rating scale analysis criteria.

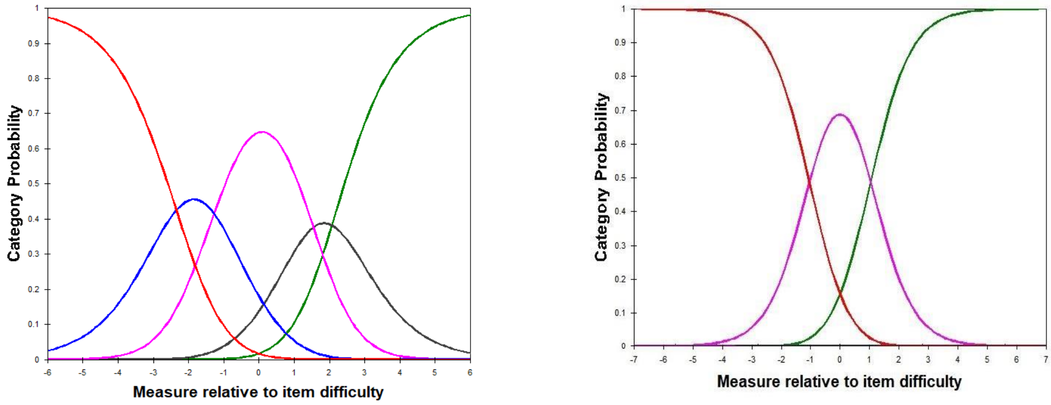

Figure 6. Probability Response to Items Graph

Based on the main findings outlined, the following recommendations are considered strategic steps to provide reinforcement and positive contributions to the development of the cognitive load scale within the context of the formal definition of limits. The unique characteristics of the mathematical context and learning environment introduce contextual variations in measurement outcomes, rendering research results non-generalizable (Fennell & Rowan, 2001; Wang et al., 2020). Referring to this, conducting external validation according to Ho et al. (2020) and Quintão et al. (2020) offers the possibility of generalizing results, validating contextual applicability, and identifying variability. Furthermore, expanding contextual variables is also a consideration in constructing the cognitive load scale. Rationalizing research findings can provide a more realistic depiction of material complexity, yield a more holistic analysis, and provide a comprehensive insight into factors contributing to cognitive load, determining which factors have the most significant impact (Breves & Stein, 2023; Mangaroska et al., 2022). Other recommendations focus on the coverage and diversity of respondents, which should at least be a researcher's consideration to ensure research results are more relevant, reduce potential bias, help improve accuracy in understanding population variability, and broaden the scope of respondents who need further empowerment (Johnson et al., 2020; Lakens, 2022).

4. CONCLUSION

In this study, we employed a Rasch model approach to examine the validity of the cognitive load instrument focused on the formal definition of limits. The results of the research indicate that the instrument is valid and capable of measuring differences in students' cognitive load. These findings provide essential insights into how the understanding of the limit concept is related to cognitive load. The results of this study have significant implications for the development of more effective mathematics teaching strategies. However, it should be noted that this research has limitations in terms of subject coverage and context. Some limitations of this study that need to be considered include: 1) the context of the cognitive load instrument is focused on students' difficulties, namely the formal definition of the limit, additional studies could explore on other specific topics in mathematics or related topics in Calculus that are notoriously known for demanding a high level of cognitive load from student; 2) the cognitive load experienced by students in the context of the formal definition of the limit is measured through the cognitive load scale, but a more in-depth examination of the general representation of the complexity level of this material is needed by involving a larger number of respondents; 3) the variability of respondents is not fully considered in the research, so differences in mathematical abilities and previous experiences are not the focus of discussion; and 4) the selection of research respondent samples is based on first-year students who are currently studying the limit function chapter, so the results are not specific to students in a particular study program. Therefore, further research is needed to confirm these findings in various educational contexts. Some recommendations that can be proposed based on the limitations of the study include conducting external validation in different mathematical contexts, expanding the context of variables to detect the complexity of material that may have a greater impact on cognitive load, considering individual factors that may affect cognitive load, and broadening the coverage and diversity of respondents. Aside from this, the Cognitive Load Scale can be useful for researchers in comparing and ordering topics in mathematics according to their cognitive load requirement. Results from these kinds of studies have implications on the way mathematics curricula are designed in such a way that topics with high cognitive load requirement are distributed and spaced out to facilitate better learning among students.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude for the financial support provided by Fundamental Research Scheme under Ministry of Education, Culture, Research, and Technology. This research was made possible and played a crucial role in the successful completion of this study. The generous funding greatly contributed to the execution of experiments, data collection, and the overall advancement of scientific knowledge in the field of Mathematics Education. The authors acknowledge the invaluable support and resources provided by Ministry of Education, Culture, Research, and Technology that enabled the successful accomplishment of this research endeavor.

REFERENCES

Adiredja, A. P. (2021). Students' struggles with temporal order in the limit definition: uncovering resources using knowledge in pieces. *International Journal of Mathematical Education in Science and Technology*, *52*(9), 1295-1321. https://doi.org/10.1080/0020739X.2020.1754477

Aghekyan, R. (2020). Validation of the SIEVEA instrument using the rasch analysis. *International Journal of Educational Research*, *103*, 101619. https://doi.org/10.1016/j.ijer.2020.101619

Al Ali, R., & Shehab, R. T. (2020). Psychometric properties of social perception of mathematics: Rasch model analysis. *International Education Studies*, *13*(12), 102-110. https://doi.org/10.5539/ies.v13n12p102

Alam, A. (2020). Challenges and possibilities in teaching and learning of calculus: A case study of India. *Journal for the Education of Gifted Young Scientists*, *8*(1), 407-433. https://doi.org/10.17478/jegys.660201

Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous rasch model that dispels any "threshold disorder controversy". *Educational and Psychological Measurement*, *73*(1), 78-124. https://doi.org/10.1177/0013164412450877

Andrich, D., & Pedler, P. (2019). A law of ordinal random error: The rasch measurement model and random error distributions of ordinal assessments. *Measurement*, *131*, 771-781. https://doi.org/10.1016/j.measurement.2018.08.062

Anmarkrud, Ø., Andresen, A., & Bråten, I. (2019). Cognitive load and working memory in multimedia learning: conceptual and measurement issues. *Educational Psychologist*, *54*(2), 61-83. https://doi.org/10.1080/00461520.2018.1554484

Arzarello, F., & Soldano, C. (2019). Approaching proof in the classroom through the logic of inquiry. In G. Kaiser & N. Presmeg (Eds.), *Compendium for Early Career Researchers in Mathematics Education* (pp. 221-243). Springer International Publishing. https://doi.org/10.1007/978-3-030-15636-7_10

Bishara, S. (2022). Linking cognitive load, mindfulness, and self-efficacy in college students with and without learning disabilities. *European Journal of Special Needs Education*, *37*(3), 494-510. https://doi.org/10.1080/08856257.2021.1911521

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press. https://doi.org/10.4324/9781410614575

Boone, W. J., & Staver, J. R. (2020). Understanding and utilizing item characteristic curves (ICC) to further evaluate the functioning of a scale. In W. J. Boone & J. R. Staver (Eds.), *Advances in Rasch Analyses in the Human Sciences* (pp. 65-83). Springer International Publishing. https://doi.org/10.1007/978-3-030-43420-5_6

Breves, P., & Stein, J.-P. (2023). Cognitive load in immersive media settings: the role of spatial presence and cybersickness. *Virtual Reality*, *27*(2), 1077-1089. https://doi.org/10.1007/s10055-022-00697-5

Brown, J. R. (2022). Rigour and thought experiments: Burgess and norton. *Axiomathes*, *32*(1), 7-28. https://doi.org/10.1007/s10516-021-09567-2

Brzezińska, J. (2020). Item response theory models in the measurement theory. *Communications in Statistics - Simulation and Computation*, *49*(12), 3299-3313. https://doi.org/10.1080/03610918.2018.1546399

Can, Y. S., Arnrich, B., & Ersoy, C. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of Biomedical Informatics*, *92*, 103139. https://doi.org/10.1016/j.jbi.2019.103139

Casale, G., Herzog, M., & Volpe, R. J. (2023). Measurement efficiency of a teacher rating scale to screen for students at risk for social, emotional, and behavioral problems. *Journal of Intelligence*, *11*(3), 57. https://doi.org/10.3390/jintelligence11030057

Case, J., & Speer, N. (2021). Calculus students' deductive reasoning and strategies when working with abstract propositions and calculus theorems. *PRIMUS*, *31*(2), 184-201. https://doi.org/10.1080/10511970.2019.1660931

Chan, S.-W., Looi, C.-K., & Sumintono, B. (2021). Assessing computational thinking abilities among Singapore secondary students: a rasch model measurement analysis. *Journal of Computers in Education*, *8*(2), 213-236. https://doi.org/10.1007/s40692-020-00177-2

Chen, C., Kang, J. M., Sonnert, G., & Sadler, P. M. (2021). High school calculus and computer science course taking as predictors of success in introductory college computer science. *ACM Trans. Comput. Educ.*, *21*(1), Article 6. https://doi.org/10.1145/3433169

Chew, S. L., & Cerbin, W. J. (2021). The cognitive challenges of effective teaching. *The Journal of Economic Education*, *52*(1), 17-40. https://doi.org/10.1080/00220485.2020.1845266

Chi, S., Liu, X., & Wang, Z. (2021). Comparing student science performance between hands-on and traditional item types: A many-facet Rasch analysis. *Studies in Educational Evaluation*, *70*, 100998. https://doi.org/10.1016/j.stueduc.2021.100998

Chi, S., Wang, Z., & Zhu, Y. (2023). Using rasch analysis to assess students' learning progression in stability and change across middle school grades. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education* (pp. 265-289). Springer International Publishing. https://doi.org/10.1007/978-3-031-28776-3_11

Chung, S., & Cai, L. (2021). Cross-classified random effects modeling for moderated item calibration. *Journal of Educational and Behavioral Statistics*, *46*(6), 651-681. https://doi.org/10.3102/1076998620983908

De Ayala, R. J. (2018). Item response theory and Rasch modeling. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd ed., pp. 145-163). Routledge. https://doi.org/10.4324/9781315755649

Eckes, T., & Jin, K.-Y. (2021). Measuring rater centrality effects in writing assessment: A bayesian facets modeling approach. *Psychological Test and Assessment Modeling*, *63*(1), 65-94.

Falotico, R., & Quatto, P. (2015). Fleiss' kappa statistic without paradoxes. *Quality & Quantity*, *49*(2), 463-470. https://doi.org/10.1007/s11135-014-0003-1

Faradillah, A., & Febriani, L. (2021). Mathematical trauma students' junior high school based on grade and gender. *Infinity Journal*, *10*(1), 53-68. https://doi.org/10.22460/infinity.v10i1.p53-68

Fennell, F., & Rowan, T. (2001). Representation: An important process for teaching and learning mathematics. *Teaching Children Mathematics*, *7*(5), 288-292. https://doi.org/10.5951/tcm.7.5.0288

Forsberg, A., Adams, E. J., & Cowan, N. (2021). Chapter one - The role of working memory in long-term learning: Implications for childhood development. In K. D. Federmeier (Ed.), *Psychology of Learning and Motivation* (Vol. 74, pp. 1-45). Academic Press. https://doi.org/10.1016/bs.plm.2021.02.001

Ghedamsi, I., & Lecorre, T. (2021). Transition from high school to university calculus: a study of connection. *ZDM – Mathematics Education*, *53*(3), 563-575. https://doi.org/10.1007/s11858-021-01262-1

Gupta, U., & Zheng, R. Z. (2020). Cognitive load in solving mathematics problems: Validating the role of motivation and the interaction among prior knowledge, worked examples, and task difficulty. *European Journal of STEM Education*, *5*(1), 5. https://doi.org/10.20897/ejsteme/9252

Gwet, K. L. (2021). Large-sample variance of fleiss generalized kappa. *Educational and Psychological Measurement*, *81*(4), 781-790. https://doi.org/10.1177/0013164420973080

Hadie, S. N. H., & Yusoff, M. S. B. (2016). Assessing the validity of the cognitive load scale in a problem-based learning setting. *Journal of Taibah University Medical Sciences*, *11*(3), 194-202. https://doi.org/10.1016/j.jtumed.2016.04.001

Hadžibajramović, E., Schaufeli, W., & De Witte, H. (2020). A rasch analysis of the burnout assessment Tool (BAT). *PLoS One*, *15*(11), e0242241. https://doi.org/10.1371/journal.pone.0242241

Hagell, P. (2019). Measuring activities of daily living in Parkinson's disease: On a road to nowhere and back again? *Measurement*, *132*, 109-124. https://doi.org/10.1016/j.measurement.2018.09.050

Ho, S. Y., Phua, K., Wong, L., & Goh, W. W. B. (2020). Extensions of the external validation for checking learned model interpretability and generalizability. *Patterns*, *1*(8). https://doi.org/10.1016/j.patter.2020.100129

Hoijtink, H. (2005). Item response models for nonmonotone items. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 373-378). Elsevier. https://doi.org/10.1016/B0-12-369398-5/00464-3

Huckaby, L. V., Cyr, A. R., Handzel, R. M., Littleton, E. B., Crist, L. R., Luketich, J. D., Lee, K. K., & Dhupar, R. (2022). Postprocedural cognitive load measurement with immediate feedback to guide curriculum development. *The Annals of Thoracic Surgery*, *113*(4), 1370-1377. https://doi.org/10.1016/j.athoracsur.2021.05.086

Indihadi, D., Suryana, D., & Ahmad, A. B. (2022). The analysis of construct validity of Indonesian creativity scale using rasch model. *Creativity Studies*, *15*(2), 560–576. https://doi.org/10.3846/cs.2022.15182

Jablonka, E. (2020). Critical thinking in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 159-163). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0_35

Jiang, D., & Kalyuga, S. (2020). Confirmatory factor analysis of cognitive load ratings supports a two-factor model. *The Quantitative Methods for Psychology*, *16*(3), 216-225. https://doi.org/10.20982/tqmp.16.3.p216

Johnson, J. L., Adkins, D., & Chauvin, S. (2020). A review of the quality indicators of rigor in qualitative research. *American Journal of Pharmaceutical Education*, *84*(1), 7120. https://doi.org/10.5688/ajpe7120

Josa, I., & Aguado, A. (2020). Measuring unidimensional inequality: Practical framework for the choice of an appropriate measure. *Social Indicators Research*, *149*(2), 541-570. https://doi.org/10.1007/s11205-020-02268-0

Katona, J. (2022). Measuring cognition load using eye-tracking parameters based on algorithm description tools. *Sensors*, *22*(3), 912. https://doi.org/10.3390/s22030912

Kidron, I. (2020). Calculus teaching and learning. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 87-94). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0_18

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in psychology*, *8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, *48*(1), 45-77. https://doi.org/10.1007/s11251-020-09502-9

Koskey, K. L., Mudrey, R. R., & Ahmed, W. (2017). Rasch derived teachers' emotions questionnaire. *Journal of Applied Measurement*, *18*(1), 67-86.

Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review*, *35*(1), 9. https://doi.org/10.1007/s10648-023-09738-0

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. https://doi.org/10.2307/2529310

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058-1072. https://doi.org/10.3758/s13428-013-0334-1

Ludyga, S., Gerber, M., & Kamijo, K. (2022). Exercise types and working memory components during development. *Trends in Cognitive Sciences*, *26*(3), 191-203. https://doi.org/10.1016/j.tics.2021.12.004

Mangaroska, K., Sharma, K., Gašević, D., & Giannakos, M. (2022). Exploring students' cognitive and affective states during problem solving through multimodal data: Lessons learned from a programming activity. *Journal of Computer Assisted Learning*, *38*(1), 40-59. https://doi.org/10.1111/jcal.12590

Martínez-Planell, R., & Trigueros, M. (2021). Multivariable calculus results in different countries. *ZDM – Mathematics Education*, *53*(3), 695-707. https://doi.org/10.1007/s11858-021-01233-6

Mutiawani, V., Athaya, A. M., Saputra, K., & Subianto, M. (2022). Implementing item response theory (IRT) method in quiz assessment system. *TEM Journal*, *11*(1), 210-218. https://doi.org/10.18421/TEM111-26

Naar, S., Chapman, J., Cunningham, P. B., Ellis, D., MacDonell, K., & Todd, L. (2021). Development of the motivational interviewing coach rating scale (MI-CRS) for health equity implementation contexts. *Health Psychology*, *40*(7), 439-449. https://doi.org/10.1037/hea0001064

Nima, A. A., Cloninger, K. M., Persson, B. N., Sikström, S., & Garcia, D. (2020). Validation of subjective well-being measures using item response theory. *Frontiers in psychology*, *10*, 3036. https://doi.org/10.3389/fpsyg.2019.03036

Oktaviyanthi, R., & Agus, R. N. (2023). Evaluating graphing quadratic worksheet on visual thinking classification: A confirmatory analysis. *Infinity Journal*, *12*(2), 207-224. https://doi.org/10.22460/infinity.v12i2.p207-224

Oktaviyanthi, R., Herman, T., & Dahlan, J. A. (2018). How does pre-service mathematics teacher prove the limit of a function by formal definition? *Journal on Mathematics Education*, *9*(2), 195-212. https://doi.org/10.22342/jme.9.2.5684.195-212

Ouwehand, K., Kroef, A. v. d., Wong, J., & Paas, F. (2021). Measuring cognitive load: Are there more valid alternatives to Likert rating scales? *Frontiers in Education*, *6*, 702616. https://doi.org/10.3389/feduc.2021.702616

Parr, E. D. (2023). Undergraduate students' interpretations of expressions from calculus statements within the graphical register. *Mathematical thinking and learning*, *25*(2), 177-207. https://doi.org/10.1080/10986065.2021.1943608

Pradipta, T. R., Perbowo, K. S., Nafis, A., Miatun, A., & Johnston-Wilder, S. (2021). Marginal region mathematics teachers' perception of using ict media. *Infinity Journal*, *10*(1), 133-148. https://doi.org/10.22460/infinity.v10i1.p133-148

Qu, Y., Kne, L., Graham, S., Watkins, E., & Morris, K. (2023). A latent scale model to minimize subjectivity in the analysis of visual rating data for the National Turfgrass Evaluation Program. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1135918

Quarfoot, D., & Rabin, J. M. (2022). A hypothesis framework for students' difficulties with proof by contradiction. *International Journal of Research in Undergraduate Mathematics Education*, *8*(3), 490-520. https://doi.org/10.1007/s40753-021-00150-z

Quintão, C., Andrade, P., & Almeida, F. (2020). How to improve the validity and reliability of a case study approach? *Journal of Interdisciplinary Studies in Education*, *9*(2), 264-275. https://doi.org/10.32674/jise.v9i2.2026

Ramakrishnan, P., Balasingam, B., & Biondi, F. (2021). Chapter 2 - cognitive load estimation for adaptive human–machine system automation. In D. Zhang & B. Wei (Eds.), *Learning Control* (pp. 35-58). Elsevier. https://doi.org/10.1016/B978-0-12-822314-7.00007-9

Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: a systematic literature review. *Artificial Intelligence Review*, *55*(3), 2495-2527. https://doi.org/10.1007/s10462-021-10068-2

Sepp, S., Howard, S. J., Tindall-Ford, S., Agostinho, S., & Paas, F. (2019). Cognitive load theory and human movement: Towards an integrated model of working memory. *Educational Psychology Review*, *31*(2), 293-317. https://doi.org/10.1007/s10648-019-09461-9

Shi, Q., Wind, S. A., & Lakin, J. M. (2023). Exploring the influence of item characteristics in a spatial reasoning task. *Journal of Intelligence*, *11*(8), 152. https://doi.org/10.3390/jintelligence11080152

Silvia, P. J., Rodriguez, R. M., Beaty, R. E., Frith, E., Kaufman, J. C., Loprinzi, P., & Reiter-Palmon, R. (2021). Measuring everyday creativity: A rasch model analysis of the biographical inventory of creative behaviors (BICB) scale. *Thinking Skills and Creativity*, *39*, 100797. https://doi.org/10.1016/j.tsc.2021.100797

Skulmowski, A. (2023). Guidelines for choosing cognitive load measures in perceptually rich environments. *Mind, Brain, and Education*, *17*(1), 20-28. https://doi.org/10.1111/mbe.12342

Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review*, *34*(1), 171-196. https://doi.org/10.1007/s10648-021-09624-7

Slavíčková, M., & Vargová, M. (2023). Differences in the comprehension of the limit concept between prospective mathematics teachers and managerial mathematicians during online teaching. In *4th International Conference, Higher Education Learning Methodologies and Technologies Online* (pp. 168-183). Cham https://doi.org/10.1007/978-3-031-29800-4_13

Stenner, A. J., Fisher, W. P., Stone, M. H., & Burdick, D. (2023). Causal rasch models. In W. P. Fisher Jr & P. J. Massengill (Eds.), *Explanatory models, unit standards, and personalized learning in educational measurement* (pp. 223-250). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-3747-7_18

Swain, T. A., Snyder, S. W., McGwin, G., Huisingh, C. E., Seder, T., & Owsley, C. (2023). Older drivers' attitudes and preferences about instrument cluster designs in vehicles revealed by the dashboard questionnaire. *Cognition, Technology & Work*, *25*(1), 65-74. https://doi.org/10.1007/s10111-022-00710-6

Sweller, J. (2011). Chapter two - Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Szulewski, A., Howes, D., van Merriënboer, J. J. G., & Sweller, J. (2021). From theory to practice: The application of cognitive load theory to the practice of medicine. *Academic Medicine*, *96*(1), 24-30. https://doi.org/10.1097/acm.0000000000003524

Tesio, L., Caronni, A., Kumbhare, D., & Scarano, S. (2023). Interpreting results from rasch analysis 1. The "most likely" measures coming from the model. *Disability and Rehabilitation*, 1-13. https://doi.org/10.1080/09638288.2023.2169771

Thompson, P. W., & Harel, G. (2021). Ideas foundational to calculus learning and their links to students' difficulties. *ZDM – Mathematics Education*, *53*(3), 507-519. https://doi.org/10.1007/s11858-021-01270-1

Toland, M. D., Li, C., Kodet, J., & Reese, R. J. (2021). Psychometric properties of the outcome rating scale: An item response theory analysis. *Measurement and Evaluation in Counseling and Development*, *54*(2), 90-105. https://doi.org/10.1080/07481756.2020.1745647

Viirman, O., Vivier, L., & Monaghan, J. (2022). The limit notion at three educational levels in three countries. *International Journal of Research in Undergraduate Mathematics Education*, *8*(2), 222-244. https://doi.org/10.1007/s40753-022-00181-0

Wang, M.-T., Degol, J. L., Amemiya, J., Parr, A., & Guo, J. (2020). Classroom climate and children's academic and psychological wellbeing: A systematic review and meta-analysis. *Developmental Review*, *57*, 100912. https://doi.org/10.1016/j.dr.2020.100912

Wu, M., Tam, H. P., & Jen, T.-H. (2016). Two-parameter IRT models. In M. Wu, H. P. Tam, & T.-H. Jen (Eds.), *Educational measurement for applied researchers: Theory into practice* (pp. 187-205). Springer Singapore. https://doi.org/10.1007/978-981-10-3302-5_10

Yamashita, T. (2022). Analyzing likert scale surveys with rasch models. *Research Methods in Applied Linguistics*, *1*(3), 100022. https://doi.org/10.1016/j.rmal.2022.100022

Yan, X., Marmur, O., & Zazkis, R. (2020). Calculus for teachers: Perspectives and considerations of mathematicians. *Canadian Journal of Science, Mathematics and Technology Education*, *20*(2), 355-374. https://doi.org/10.1007/s42330-020-00090-x