Jurnal Penelitian dan Evaluasi Pendidikan
Volume 25, No 2, December 2021 (126-137)
Online: http://journal.uny.ac.id/index.php/jpep

Validity of critical thinking skills instrument on prospective Mathematics teachers

Ayu Faradillah*; Sabila Adlina
Universitas Muhammadiyah Prof. Dr. HAMKA, Indonesia
*Corresponding Author. E-mail: ayufaradillah@uhamka.ac.id

ARTICLE INFO

Article History
Submitted:
8 May 2021
Revised:
13 October 2021
Accepted:
27 October 2021

Keywords
critical thinking skills; Rasch model; CFA

Scan Me:

ABSTRACT

This study aims to describe the process of validity of critical thinking skills on prospective mathematics teachers. This research used a quantitative approach with a survey method. Data were collected from 245 prospective mathematics teachers from 19 Mathematics education study programs at 19 higher education institutions in Indonesia. The data were collected using a questionnaire given via Google Form and analyzed by the Rasch Model analysis using Winstep software and Confirmatory Factor Analysis (CFA) using JASP. The results show that the instrument of critical thinking skills with indicators such as open-mindedness, inquisitiveness, systematicity, truth-seeking, analyticity, and self-confidence is valid and reliable, although it has to consider eliminating items and person misfit.

How to cite:
Faradillah, A., & Adlina, S. (2021). Validity of critical thinking skills instrument on prospective Mathematics teachers. *Jurnal Penelitian dan Evaluasi Pendidikan, 25*(2), 126-137. doi:https://doi.org/10.21831/pep.v25i2.40662

INTRODUCTION

The ability to think critically is one of the goals that is recognized at the educational level (Sustekova et al., 2019). In Indonesia, the ability to think critically is also one of the main goals in education. Critical thinking skills are needed in education because with the ability to think critically, students can prepare themselves for future conditions (Nadeak et al., 2020). Indonesia needs its students to be equipped with competency, skills, and future builders characteristics, including critical thinking skills, communication skills, good attitudes, teamwork skills, and civic intelligence (Senk & Thompson, 2020). In learning, critical thinking skills have become the requirement that must be fulfilled (Pradana et al., 2017), one of which is in learning mathematics. However, this ability is used in learning mathematics and is also applied in everyday life. The ability to think critically in mathematics requires not only comprehension and knowledge in reaching the right solution, but also understanding, explaining, investigating various ways of finding solutions, and reflecting the benefits of mathematics in everyday life (Dolapcioglu & Doğanay, 2020).

This shows the importance of the ability to think critically; thus, in the process, critical thinking is seen as an intellectual process of application, analysis, synthesis, conceptualization, and evaluation of the information gathered through observation, experience, reflection, reasoning, and communication, which are found in the way action is based on universal intellectual values (Dinuţă, 2015). With the ability to think critically, we will find problems and solve them in an effective way of thinking.

10.21831/pep.v25i2.40662

Therefore, many researchers have researched critical thinking skills and conducted STEM research on critical thinking models to obtain results through STEM-based activities that significantly play an important role in developing critical thinking skills (Mater et al., 2020). Furthermore, Yolanda, who researched problem-based learning on critical thinking skills, shows that the PBL approach can significantly influence critical thinking skills (Yolanda, 2019). Both studies used different critical thinking skills instruments. This shows that there is currently no appropriate instrument to measure critical thinking skills. Hence, this becomes one of the challenges for researchers in which there is a limited measurement of valid or reliable critical thinking skills instruments (Quinn et al., 2020). With valid and reliable instruments, it shows that the instrument is appropriate to measure what must be measured and can produce the same data if we measure the same object repeatedly. Thus, using valid and reliable instruments will produce valid and reliable research results (Sugiyono, 2016). Based on this problem, solutions and innovations are needed to find the right instrument, which can be done by instrument validity. By doing the validity, we can determine the accuracy to what extent the instrument can be measured, and this is relevant to previous researches (Mapeala & Siew, 2015; Purnami et al., 2021; Quinn et al., 2020; Sustekova et al., 2019).

A research by Mapeala and Siew (2015) revealed that the Test of Science Critical Thinking Test (TSCT) as a critical thinking test instrument conducted on fifth-grade students is valid. Purnami et al. (2021) research revealed the Scientific Group Inquiry Learning (SGIL) model as an instrument to measure critical thinking skills given to PGSD students was valid. Research by Quinn et al. (2020) shows that the Student Educator Negotiated Critical Thinking Dispositions Scale (SENCTSD) is a valid measure of critical thinking disposition. Furthermore, Sustekova et al. (2019) revealed that the Slovak version of the critical thinking test instrument tested on students of the Faculty of Humanities could be validated.

Research conducted by Mapeala and Siew (2015), Quinn et al. (2020), and Sustekova et al. (2019) discussed the same variable, namely the validity of critical thinking skills. The instrument for measuring critical thinking skills used by Mapeala and Siew (2015) is used in learning physics. Purnami et al. (2021) use an instrument to measure critical thinking skills in learning basic science concepts. The relevant research has the same area of application of critical thinking skills in science learning, indicating that the critical thinking skills instrument is currently underutilized in mathematics learning. In Mapeala and Siew (2015) and Sustekova et al. (2019), they used different respondents, namely fifth-grade students and students of the Faculty of Humanities. Based on the explanation, there is a gap, namely different respondents, so the researchers used students as respondents in this study. The novelty in this study is that the students used as respondents were prospective mathematics teachers. From the relevant research described, no one has conducted research on the validity of the critical thinking skills instrument and prospective mathematics teachers. This study uses a critical thinking skills instrument adapted from the California Critical Thinking Dispositions Inventory (CCDTI) with the indicators of Open-Mindedness, Inquisitiveness, Systematicity, Truth-Seeking, Analyticity, and Self Confidence. This adaptation research is used to determine whether there are differences in results in the use of instruments carried out in different regions, such as research conducted by Sustekova, who conducted adaptation research on the HCTA critical thinking skills instrument, which was adapted to the conditions of the Slovakia region (Sustekova et al., 2019), and adaptation research can facilitate the process of instrument validity and reliability.

RESEARCH METHOD

This study used a quantitative approach with a survey method. The quantitative approach is done by asking the same questions to all respondents, so the researchers have data that can be compared across all existing samples (Mat Roni et al., 2020). The survey method used is a validated survey modified according to the research objectives and collected data

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.40662

(Faradillah & Febriani, 2021; Mat Roni et al., 2020). This study aims to describe the process of the validity of the critical thinking skills instrument on prospective mathematics teachers. The subjects of this study were 245 prospective mathematics teachers in Indonesia, consisting of 209 female and 36 male students. Data collection was carried out using a questionnaire given online via a Google Form. The instrument used was a critical thinking skills questionnaire. The instrument has gone through validation by experts, namely two lecturers of Mathematics Education and English Language Education. The two validators stated that the questionnaire was suitable for use with improvement, namely the use of language that was easy to understand by students. The critical thinking skills questionnaire contains responses using a Likert scale of 1-6 with statements, which strongly disagree, disagree, somewhat disagree, somewhat agree, agree, strongly agree. This critical thinking skills instrument was adapted from previous research and has been adapted to a language structure that is easy to understand (Yorgancı, 2016). The indicators of critical thinking skills are presented in Table 1.

Table 1. The Indicators of Critical Thinking

No	Indicators of Critical Thinking	Statement Item Number	Total
1	Open-Mindedness	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	12
2	Inquisitiveness	13, 14, 15, 16, 17, 18, 19, 20	8
3	Systematicity	21, 22, 23, 24, 25, 26	6
4	Truth-Seeking	27, 28, 29, 30, 31, 32, 33	7
5	Analyticity	34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44	11
6	Self Confidence	45, 46, 47, 48, 49, 50, 51	7

The Rasch model employing Winstep software was used to analyze the data. The data obtained were processed from raw scores to logit interval values using the Rasch model in the form of Ms. Excel (Hadi & Faradillah, 2019). The results obtained are students' responses based on the questionnaire given in the form of a Wright maps table. Wright maps are a technique for displaying data sequences based on the difficulty level of the question item (Boone, 2016). Confirmatory Factor Analysis (CFA) was done using JASP. CFA is used to measure construct validity (Abraham et al., 2020). Data analysis using CFA is seen from loading factor criteria >0.3 – >0.4 and the model fit or goodness of fit based on Chi-Square X^2, p-value, RMSEA, SRMR, GFI, TLI, and CFI (Darodjat & Zuchdi, 2016; Hair Jr. et al., 2018; Sugiyono, 2016; Yuniarti & Soenarto, 2016). The criteria for fit models are presented in Table 2.

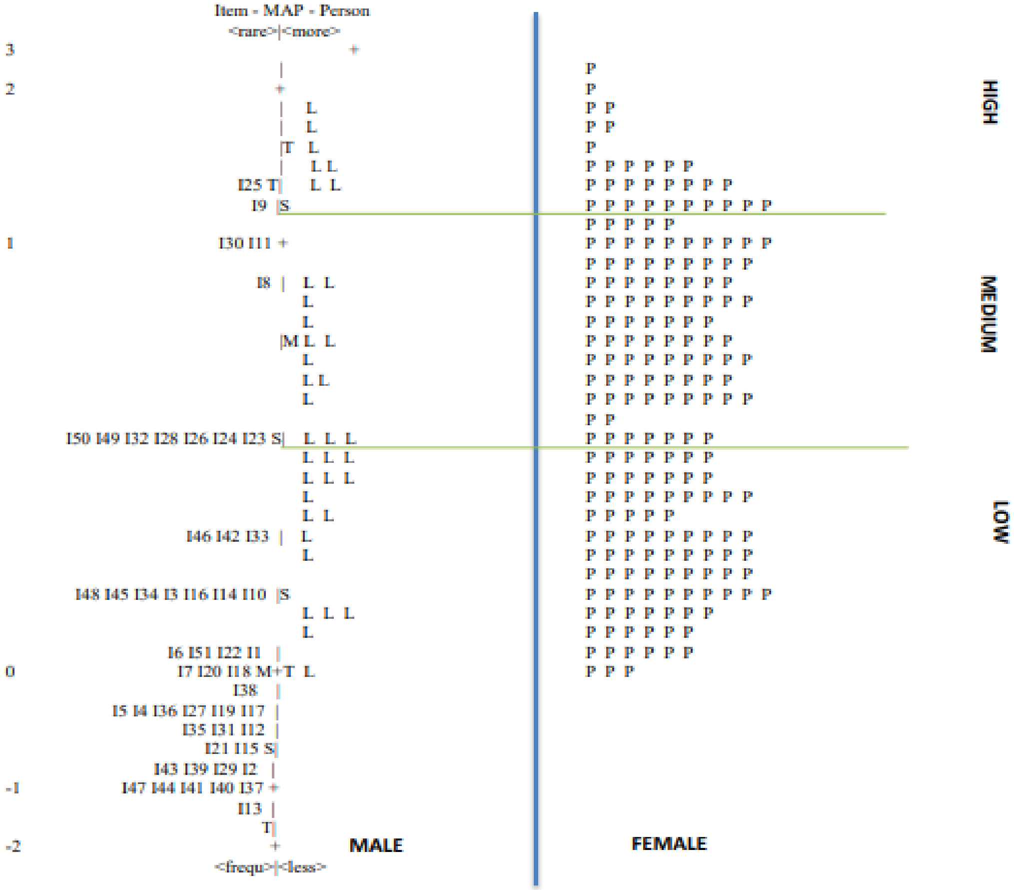
Table 2. Fit Model Indices

Statistics Fit	Criteria
X^2 Chi-Square	Expected Small
p-values	≥0.05
RMSEA	≤0.08
SRMR	≤0.09
GFI	0.80≤GFI≤0.90
CFI	≥0.90
TLI	≥0.90

This research is viewed by gender. In Figure 1, the Wright maps show the item and person plots. The person plot shows good test performance if the person size is high. The person plot shows the size from the highest to the lowest (Boone, 2016). It can be seen that 38 (15.5%) students have high critical thinking skills, 104 (42.4%) students have moderate critical thinking skills, and 103 (42.0%) students have low critical thinking skills. The Rasch model is also used to determine the validity and reliability of the critical thinking skills questionnaire and to determine the validity and reliability of the data used to evaluate the Cronbach's Alpha (Osman et al., 2016).

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.40662



Annotation: L = male P = female

Figure 1. The Wright Maps of Critical Thinking Categories based on Gender

Table 3. Reliability in Rasch Analysis (Sumintono & Widhiarso, 2014)

Statistics	Fit Indices	Interpretation
Cronbach's Alpha (KR-20)	<0.5	Low
	0.5-0.6	Moderate
	0.6-0.7	Good
	0.7-0.8	High
	>0.8	Very High
Item and Person Reliability	<0.67	Low
	0.67-0.80	Sufficient
	0.81-0.90	Good
	0.91-0.94	Very Good
	>0.94	Excellent
Item and Person Separation		A High separation value indicates that the quality of the instrument is good because it can identify groups of items and a person

Table 4. The Summary of Instrument Statistics

	Mean	Separation	Reliability	Cronbach's α
Person	0.02	2.04	0.81	0.84
Item	0.09	8.47	0.99	

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.40662

In Table 3, the reliability of the Rasch Model consists of three criteria for the fit index, namely Cronbach's alpha, item, and person reliability, and also item and person separation (Sumintono & Widhiarso, 2014). Meanwhile, Table 4 shows a summary of the statistical instruments, which consist of the reliability item and person, the separation item and person, the mean item and person, as well as Cronbach's alpha. Item and person reliability show reliable measurement results. The coefficient of Cronbach's alpha is 0.84, which is categorized as very high. The reliability result of the item with a value of 0.99 shows that the item that is used has a very good reliability value, and the separation item shows a score of 8.47, which indicates that it is divided into nine groups of items (8.74 roundings).

Table 5. Fit Indices for Item Fit and Person Fit (Sumintono & Widhiarso, 2014)

Statistics	Fit Indices
Outfit Mean Square Values (MNSQ)	0.5 – 1.5
Outfit Z-Standard Values (ZSTD)	-2.0 - +2.0
Point Measure Correlation (PTMEA-CORR)	0.4 – 0.85

In Table 5, there are Outfit Mean Square Values (MNSQ), Outfit Z-Standard Values (ZSTD), and also Point Measure Correlation (PTMEA-CORR) which are the criteria for assessing the suitability of the item and person. Item fit can show that the item functions appropriately according to the size it should be, but if the item does not fit, or misfit, it shows that the respondent has misconceptions about the item (Faradillah & Febriani, 2021). On the other hand, for person fit, it can also be seen from the three criteria (Sumintono & Widhiarso, 2014).

FINDINGS AND DISCUSSION

After 51 statement items were running, it turned out that the validity was 34.2%. Based on the requirements, raw variance data has a minimum of 20%, if the data is more than 40%, it means good, and if the data is more than 60%, it means excellent (Sumintono & Widhiarso, 2014). In Table 6, it is less than 60%, so the validity was low. Hence, it is necessary to retest by looking at items and people that are misfits. Misfit items and persons were removed, then the data was retested again.

This study aims to describe the process of the validity of the critical thinking skills instrument on prospective mathematics teachers. The analysis results that are carried out descriptively can be seen using the Rasch model analysis. Validity can be seen from the item and person fit that has been run twice for an item and four times for a person. This is conducted to get the data that fits the criteria. The Rasch model analysis can perform an analysis process that is carried out repeatedly until it gets data that fits the criteria (Sumintono & Widhiarso, 2014).

Table 6. Standardized Residual Variance (in Eigenvalue Units)

	Empirical		Modeled
Raw Variance Explained by Measure	24.1	32.1%	34.2%

Item Fit

Item fit can show that the item functions appropriately according to the size it should be, but if the item does not fit or misfit, it shows that the respondent has misconceptions about the item (Faradillah & Febriani, 2021). The missing items can be seen from the three criteria, namely MNSQ Outfit, ZSTD Outfit, and also PTMEA-CORR. The data which do not meet the criteria need to be removed or corrected. This is done because it does not fit the

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.40662

model (Tabatabaee-Yazdi et al., 2018). Therefore, in this study, the item fit data in Table 7 were obtained after running the data twice, because in the first and second running data, there were still data that did not meet the criteria.

Table 7. Misfit Order of the Items

Person	Measure	Outfit MNSQ (0.5 - 1.5)	Outfit ZSTD (-2.0 - +2.0)	PTMEA-CORR (0.4-0.85)
I3	0.28	2.20	9.70	0.18
I32	0.56	1.78	7.60	0.18
I5	-0.34	1.68	5.50	0.20
I1	0.08	1.62	5.40	0.09
I2	-0.72	1.58	4.80	0.30
I11	1.01	1.55	6.30	0.02
I25	1.35	1.52	5.80	0.20
I12	-0.45	1.62	5.10	0.26
I36	-0.31	1.58	4.80	0.36
I10	0.39	1.57	5.10	0.17
I34	0.36	1.44	4.10	0.23
I33	0.64	1.40	4.00	0.21
I28	0.77	1.40	4.10	0.14
I9	1.48	1.38	4.30	0.04
I27	-0.34	1.37	3.30	0.36
I6	0.12	1.36	3.30	0.34
I30	1.23	1.36	4.10	0.02
I8	1.14	1.34	3.80	0.17
I24	0.77	1.25	2.70	0.23
I38	-0.11	1.23	2.10	0.37
I22	0.14	1.23	2.10	0.31
I4	-0.22	2.03	6.80	0.41
I44	-1.31	1.53	4.10	0.46
I26	1.21	1.59	4.80	0.19
I23	1.33	1.30	2.60	0.35

Table 7 presents the items that are not suitable or misfit, which means that these items do not meet the criteria of MNSQ Outfit, ZSTD Outfit, and PTMEA-CORR. There were 25 misfit statement items out of the 51 statement items tested. There were 11 statement items misfit because they did not meet the three recommended criteria, and 14 statements were misfit because they did not meet the two recommended criteria (Sumintono & Widhiarso, 2014). In addition, there were 13 MNSQ Outfit items more than 1.4, so that these items need to be removed (Mohamad et al., 2015). For item fit, there are 25 misfit statement items because they do not meet the criteria of the MNSQ Outfit, ZSTD Outfit, and PTMEA-CORR, therefore, these items need to be removed. Meanwhile, 26 statement items match the recommended criteria (Sumintono & Widhiarso, 2014), which means that these items have good quality.

Person Fit

Rasch model analysis is used in order to determine person fit. Person fit can be seen with the Rasch model that is seen from the unusual students' response patterns (Faradillah & Febriani, 2021). Inappropriate or misfit person data can be seen from MNSQ Outfit, ZSTD Outfit, and also PTMEA-CORR. The misfit data needs to be removed or corrected. In this

10.21831/pep.v25i2.40662

study, the person fit data, as presented in Table 8, were obtained after running the data four times, because in the first until the third running data, there were still data that did not meet the criteria.

Table 8. Summary Misfit Order of the Person

	Total Subject	Total Score	Measure	Outfit MNSQ		Outfit ZTSD		PTMEA-CORR	
				Below 0.5	**Above 1.5**	**Below -2.0**	**Above +2.0**	**Below 0.4**	**Above 0.85**
F	87	91-264	-0.07-1.89	13	52	36	51	23	0
M	18	99-286	0.10-2.90	3	14	3	15	8	0

Table 8 shows the responses of students who do not match the criteria suggested in the Rasch Model analysis. There were 105 (42.9%) students out of 245 students, with code F as female and M as male, indicating that their responses did not match the given criteria. There were 29 students who showed responses that did not match the three suggested criteria, and 76 students showed responses that did not match the two suggested criteria (Sumintono & Widhiarso, 2014). Thus, in this study, there were 140 (57.1%) students who showed responses according to the suggested criteria, so it can be said that the responses were of good quality. For person fit, 105 (42.9%) students out of 245 did not meet the MNSQ Outfit, ZSTD Outfit, and PTMEA-CORR criteria.

Reliability

Reliability is a measuring tool that provides the same results from an instrument. Therefore, the instrument must have the same and consistent score when used at different times (Mohamad et al., 2015; Sundayana, 2018). Reliability and analysis used the Rasch model with Winstep software. Statistics are used to measure the reliability between statement items. Higher values indicate a strong relationship between statement items, whereas lower ones indicate a weak relationship between statement items (Mohamad et al., 2015).

Table 9. The Value of Person Cronbach's Alpha (KR-20), Person Reliability, Item Reliability, Person Separation, and Item Separation

Statistics	Value
Cronbach's alpha (KR-20)	0.84
Person Reliability	0.81
Item Reliability	0.99
Person Separation	2.04
Item Separation	8.47

Table 9 shows the Cronbach's alpha (KR-20) value of 0.84. If the Cronbach's alpha value is higher than 0.8 (Table 3), it shows the meaning of "very high". The value of person reliability is 0.81, and item reliability is 0.99. Based on Table 3, item reliability with a value of 0.81-0.90 means "good" and item reliability with a value higher than 0.94 means "excellent". Based on the aforementioned explanation, the instrument in this study has a Cronbach's alpha (KR-20) score of 0.84, which means very high, a person reliability score of 0.81 which means good, and an item reliability score of 0.99 which means excellent.

Table 10. KMO, Bartlett's Test, and Total Variance Explained

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.784
Bartlett's Test of Sphericity	Approx Chi-Square	4979.151
	Sig.	0.000
Total Variance Explained		49.450%

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

Table 10 shows the KMO value of 0.784. If the KMO is more than 0.5, it can be said that the miserable or data analysis can be continued (Sadtyadi & Kartowagiran, 2014). The value of Bartlett's test of sphericity is 4979.151 with a sig. 0.0000, which mean that there is a significant correlation between the statement. KMO value and Bartlett's test of sphericity are used as requirements for further analysis (Sadtyadi & Kartowagiran, 2014). In this study, the KMO value and Bartlett's test of sphericity met the required criteria. Based on the eigenvalues greater than one and the rotated factor matrix > 0.3, it shows the total variance explained is 49.450%, meaning that the critical thinking skills variable can be measured as a whole based on 49.450% (Suranto et al., 2014).

Factor Analysis

The factor analysis that is used is Confirmatory Factor Analysis (CFA). CFA is used for construct validity. CFA is aimed at determining the construct validity of a theory based on measurements (Kumalasari et al., 2020). In order to find out the validity of the instrument, it can be seen from the loading factor value >0.3 - >0.4 that it can be declared valid (Darodjat & Zuchdi, 2016; Hair Jr. et al., 2018). In addition to looking at the loading factor value to determine validity, it is necessary to first fit the entire model, namely goodness of fit (Yuniarti & Soenarto, 2016). The criteria that are used to see the goodness of fit are p-values, chi-square (X^2 *df*) is expected to be small, RMSEA (Root Mean Square Error Of Approximation) with a value of ≤0.08, SRMR (Standardizes Root Mean Square Residual) with criteria a value of ≤0.09, GFI (Goodness of Fit Index) with a criteria value of 0.80 ≤ GFI ≤ 0.90, TLI (Tucker Lewis Index) with a value of ≥0.09, and CFI (Comparative Fit Index) with a criteria value of ≥0.09 (Darodjat & Zuchdi, 2016; Hair Jr. et al., 2018; Sugiyono, 2016; Yuniarti & Soenarto, 2016).

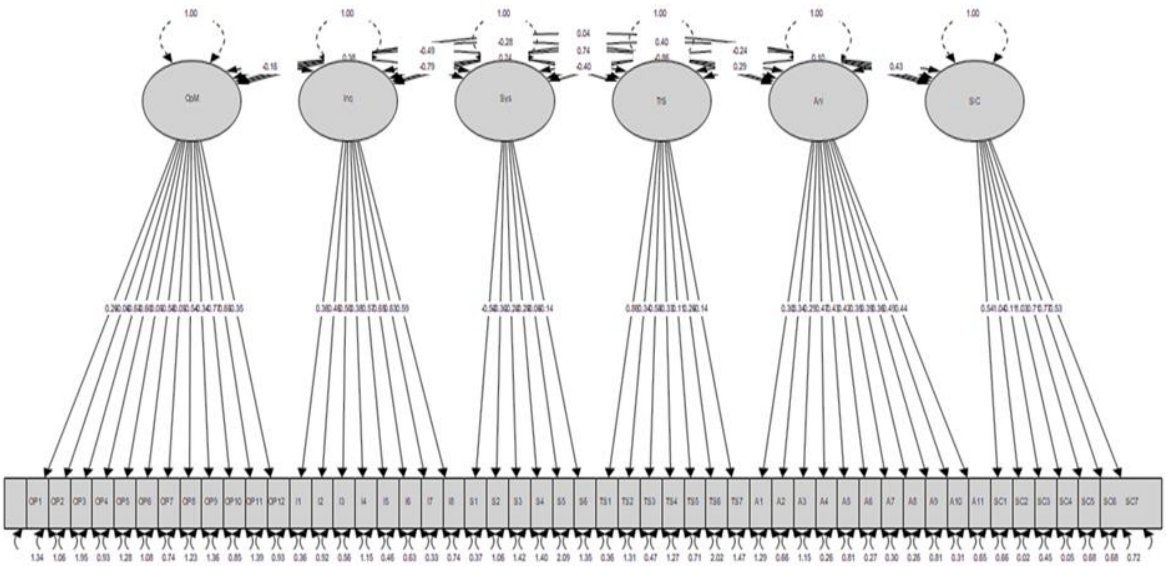


Figure 2. Confirmatory Factor Analysis of Critical Thinking Skill

Table 11. Model Fit Before Modification

Statistics Fit	Criteria	Output	Interpretation
X^2 Chi-Square	Expected Small	2788.356	Poor Fit
p-values	≥0.05	<0.01	Poor Fit
RMSEA	≤0.08	0.073	Good Fit
SRMR	≤0.09	0.100	Poor Fit

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.40662

Based on the results of the CFA analysis as presented in Figure 2, Table 11 shows that the fit model of the goodness of fit has not met the fit criteria. Therefore, it is necessary to modify the model to get fit criteria by removing the unsuitable loading factor value, which is <0.3 (Hair Jr. et al., 2018). Items with a loading factor value below the criteria show low quality (Kumalasari et al., 2020). From an analysis of the initial model, nineteen items did not meet the criteria, namely OP1, OP3, OP4, OP6, OP8, OP9, OP10, OP11, OP12, S1, S2, S3, TS4, TS5, TS6, TS7, A1, A2, A3, SC3, so they must be removed.

Table 12. Result of Factor Loading After Modification

Factor	Indicator	Symbol	Estimate	Std. Error	z-value	P	95% Confidence Interval		Std. Est. (all)
							Lower	Upper	
Open Mindedness	OP2	λ11	0.374	0.092	4.085	< .001	0.195	0.554	0.363
	OP5	λ12	0.316	0.094	3.355	< .001	0.132	0.501	0.279
	OP7	λ13	0.305	0.076	4.025	< .001	0.157	0.454	0.355
Inquistiveness	I1	λ21	0.375	0.045	8.272	< .001	0.286	0.464	0.533
	I2	λ22	0.442	0.071	6.258	< .001	0.304	0.581	0.417
	I3	λ23	0.502	0.057	8.740	< .001	0.390	0.615	0.559
	I4	λ24	0.389	0.077	5.072	< .001	0.239	0.540	0.343
	I5	λ25	0.565	0.055	10.232	< .001	0.457	0.673	0.636
	I6	λ26	0.642	0.064	10.044	< .001	0.517	0.767	0.626
	I7	λ27	0.635	0.051	12.501	< .001	0.535	0.734	0.742
	I8	λ28	0.585	0.067	8.767	< .001	0.454	0.716	0.560
Systematicity	S4	λ31	0.497	0.100	4.979	< .001	0.301	0.693	0.408
	S5	λ32	0.417	0.117	3.578	< .001	0.189	0.646	0.288
	S6	λ33	0.837	0.120	6.991	< .001	0.602	1.072	0.717
Truth-Seeking	TS1	λ41	0.922	0.084	11.017	< .001	0.758	1.086	0.879
	TS2	λ42	0.342	0.084	4.067	< .001	0.177	0.507	0.286
	TS3	λ43	0.551	0.065	8.454	< .001	0.423	0.678	0.613
Analyticity	A2	λ51	0.335	0.059	5.668	< .001	0.219	0.450	0.379
	A4	λ52	0.468	0.042	11.093	< .001	0.385	0.551	0.675
	A5	λ53	0.432	0.066	6.533	< .001	0.303	0.562	0.432
	A6	λ54	0.432	0.041	10.493	< .001	0.351	0.513	0.646
	A7	λ55	0.360	0.041	8.723	< .001	0.279	0.441	0.556
	A8	λ56	0.388	0.040	9.632	< .001	0.309	0.467	0.603
	A9	λ57	0.365	0.065	5.609	< .001	0.238	0.493	0.376
	A10	λ58	0.487	0.046	10.635	< .001	0.397	0.577	0.653
	A11	λ59	0.430	0.060	7.142	< .001	0.312	0.548	0.468
Self Confidence	SC1	λ61	0.538	0.058	9.302	< .001	0.424	0.651	0.550
	SC2	λ62	1.042	0.048	21.621	< .001	0.948	1.136	0.991
	SC4	λ63	1.026	0.049	21.004	< .001	0.931	1.122	0.976
	SC5	λ64	0.711	0.062	11.490	< .001	0.589	0.832	0.654
	SC6	λ65	0.774	0.063	12.205	< .001	0.650	0.899	0.685
	SC7	λ66	0.530	0.060	8.890	< .001	0.413	0.647	0.530

The modification is done by eliminating items with a loading factor value <0.3. Them 32 items out of 51 items meet the recommended fit criteria, namely the loading factor value >0.3 (Hair Jr. et al., 2018). The loading factor value >0.3 indicates a good quality item. After modification on the model, the fit value or goodness of fit shows the fit criteria (Table 12). Chi-Square X^2 value of 933.559 fits based on the expected criteria is small (Darodjat & Zuchdi, 2016). The p-value does not fit because the value is less than the recommended criteria of ≥0.05. It is estimated that the p-value is sensitive to the number of samples (Kumalasari et al., 2020). The p-value that is a poor fit or not suitable indicates a lack of fit in the covariance matrix, but not all criteria in the fit model must match or fit because if one of the models is fit, then the model can be said to be fit (Hair Jr. et al., 2018). The RMSEA and SRMR values indicate fit criteria with values of 0.066 and 0.081 according to the recommended criteria. RMSEA shows the most informative fit model indicators, SRMR shows a residual fit of the variance-covariance matrix of the sample data (Yuniarti & Soenarto, 2016). CFI and TLI show a value of 0.824, and 0.805 indicates a marginal fit value criterion. The GFI shows the accuracy of the model I producing a covariance matrix based on the result of the GFI analysis

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.40662

showing the fit criteria because the value is greater than the recommended 0.80 (Yuniarti & Soenarto, 2016). Based on this explanation, the critical thinking skills instrument, which was analyzed using CFA, showed a valid result.

CONCLUSION

Based on the results of the analysis carried out using the Rasch Model analysis, the 25 misfit items that did not match the criteria for Outfit Mean Square Values (MNSQ), Outfit Z-Standardized Values (ZSTD), Point Measure Correlation (PTMEA-CORR) need to be eliminated, and 105 (42.9%) of the students' responses that were misfit because they did not match the criteria need to be eliminated as well. Therefore, there were 140 (57.1%) students' responses matched, indicating the assessment's good quality. The instruments in this study were of very high quality. Meanwhile, based on the analysis using CFA, 32 items out of 51 valid items have a loading factor > 0.3 and the instrument shows a model fit index that is fit after eliminating statement items that do not match the criteria. Thus, from the results of this study, it can be concluded that the critical thinking skills instrument is valid and reliable. However, it must consider item removal and person misfit.

REFERENCES

Abraham, J., Ali, M. M., Andangsari, E. W., & Hartanti, L. E. P. (2020). Confirmatory factor analysis of celebrity worship, digital literacy, and nostalgia: Dataset of Indonesians. *Data in Brief*, *33*, 106417. https://doi.org/10.1016/j.dib.2020.106417

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education*, *15*(4), rm4. https://doi.org/10.1187/cbe.16-04-0148

Darodjat, D., & Zuchdi, D. (2016). Model evaluasi pembelajaran akidah dan akhlak di Madrasah Tsanawiyah (MTs). *Jurnal Penelitian Dan Evaluasi Pendidikan*, *20*(1), 11–26. https://doi.org/10.21831/pep.v20i1.7517

Dinuță, N. (2015). The use of critical thinking in teaching Geometric Concepts in primary school. *Procedia - Social and Behavioral Sciences*, *180*, 788–794. https://doi.org/10.1016/j.sbspro.2015.02.205

Dolapcioglu, S., & Doğanay, A. (2020). Development of critical thinking in mathematics classes via authentic learning: An action research. *International Journal of Mathematical Education in Science and Technology*, 1–24. https://doi.org/10.1080/0020739X.2020.1819573

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

Hadi, W., & Faradillah, A. (2019). The algebraic thinking process in solving HOTS questions reviewed from student achievement motivation. *Al-Jabar: Jurnal Pendidikan Matematika*, *10*(2), 327–337. https://doi.org/10.24042/ajpm.v10i2.5331

Hair Jr., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis*. Cengage. https://www.cengage.co.uk/books/9781473756540/

Kumalasari, D., Luthfiyani, A. N., & Grasiawaty, N. (2020). Analisis faktor adaptasi instrumen resiliensi akademik versi Indonesia: Pendekatan eksploratori dan konfirmatori. *JPPP - Jurnal Penelitian Dan Pengukuran Psikologi*, *9*(2), 84–95. https://doi.org/10.21009/JPPP.092.06

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.40662

Mapeala, R., & Siew, N. M. (2015). The development and validation of a test of science critical thinking for fifth graders. *SpringerPlus*, *4*(1), 741. https://doi.org/10.1186/s40064-015-1535-0

Mat Roni, S., Merga, M. K., & Morris, J. E. (2020). *Conducting quantitative research in education*. Springer Singapore. https://doi.org/10.1007/978-981-13-9132-3

Mater, N. R., Haj Hussein, M. J., Salha, S. H., Draidi, F. R., Shaqour, A. Z., Qatanani, N., & Affouneh, S. (2020). The effect of the integration of STEM on critical thinking and technology acceptance model. *Educational Studies*, 1–17. https://doi.org/10.1080/03055698.2020.1793736

Mohamad, M. M., Sulaiman, N. L., Sern, L. C., & Salleh, K. M. (2015). Measuring the validity and reliability of research instruments. *Procedia - Social and Behavioral Sciences*, *204*, 164–171. https://doi.org/10.1016/j.sbspro.2015.08.129

Nadeak, B., Juwita, C. P., Sormin, E., & Naibaho, L. (2020). Hubungan kemampuan berpikir kritis mahasiswa dengan penggunaan media sosial terhadap capaian pembelajaran pada masa pandemi Covid-19. *Jurnal Konseling Dan Pendidikan*, *8*(2), 98–104. https://doi.org/10.29210/146600

Osman, S. A., Khoiry, M. A., Rahman, N. A., Rahni, A. A. A., Mansor, M. R. A., Nordin, D., & Johar, S. (2016). The effectiveness of industrial training from the perspective of students of the Civil and Structure Engineering Department. *Journal of Engineering Science and Technology*, *Special Issue on PEKA*, 1–12. https://jestec.taylors.edu.my/Special Issue_PEKA_2016/PEKA 2016_paper 1.pdf

Pradana, S. D. S., Parno, P., & Handayanto, S. K. (2017). Pengembangan tes kemampuan berpikir kritis pada materi Optik Geometri untuk mahasiswa Fisika. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *21*(1), 51–64. https://doi.org/10.21831/pep.v21i1.13139

Purnami, W., Ashadi, A., Suranto, S., Sarwanto, S., Sumintono, B., & Wahyu, Y. (2021). Investigation of person ability and item fit instruments of eco critical thinking skills in Basic Science Concept materials for elementary pre-service teachers. *Jurnal Pendidikan IPA Indonesia*, *10*(1), 127–137. https://doi.org/10.15294/jpii.v10i1.25239

Quinn, S., Hogan, M., Dwyer, C., Finn, P., & Fogarty, E. (2020). Development and validation of the Student-Educator Negotiated Critical Thinking Dispositions Scale (SENCTDS). *Thinking Skills and Creativity*, *38*, 100710. https://doi.org/10.1016/j.tsc.2020.100710

Sadtyadi, H., & Kartowagiran, B. (2014). Pengembangan instrumen penilaian kinerja guru sekolah dasar berbasis tugas pokok dan fungsi. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(2), 290–304. https://doi.org/10.21831/pep.v18i2.2867

Senk, S. L., & Thompson, D. R. (2020). School mathematics curricula. In *Standards-based school mathematics curricula* (pp. 3–28). Routledge. https://doi.org/10.4324/9781003064275-1

Sugiyono, S. (2016). *Statistika untuk penelitian*. Alfabeta.

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial.* Trim Komunikata.

Sundayana, R. (2018). *Statistika penelitian pendidikan* (4th ed.). Alfabeta.

Suranto, S., Muhyadi, M., & Mardapi, D. (2014). Pengembangan instrumen evaluasi Uji Kompetensi Keahlian (UKK) administrasi perkantoran di SMK. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(1), 98–114. https://doi.org/10.21831/pep.v18i1.2127

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.40662

Sustekova, E., Kubiatko, M., & Usak, M. (2019). Validation of critical thinking test on Slovak conditions. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(12), em1798. https://doi.org/10.29333/ejmste/112295

Tabatabaee-Yazdi, M., Motallebzadeh, K., Ashraf, H., & Baghaei, P. (2018). Development and validation of a teacher success questionnaire using the Rasch model. *International Journal of Instruction*, *11*(2), 129–144. https://doi.org/10.12973/iji.2018.11210a

Yolanda, F. (2019). The effect of problem based learning on mathematical critical thinking skills of junior high school students. *Journal of Physics: Conference Series*, *1397*(1), 012082. https://doi.org/10.1088/1742-6596/1397/1/012082

Yorgancı, S. (2016). Critical thinking dispositions of pre-service mathematics teachers. *Participatory Educational Research*, *3*(3), 36–46. https://doi.org/10.17275/per.16.13.3.3

Yuniarti, N., & Soenarto, S. (2016). Validitas konstrak instrumen evaluasi outcome lembaga pendidikan guru vokasional. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *20*(2), 221–233. https://doi.org/10.21831/pep.v20i2.8448

ISSN (print) 2685-7111 | ISSN (online) 2338-6061