

Analyzing the Quality of Instrument for Critical Thinking Skill and Assessing Students' Critical Thinking Skill in Ecology using the Rasch Model

Tantri Tania^{a)*}, Sajidan^{1,b)*}, Harlita^{2,c)*}

¹⁾Department of Biology Teacher Training and Education, Postgraduate Program, Universitas Sebelas Maret

²⁾Department of Biology Teacher Training and Education, Postgraduate Program, Universitas Sebelas Maret

³⁾Department of Biology Education Faculty of Teacher Training and Education, Universitas Sebelas Maret

^{a)}Corresponding author: tantritania2016@gmail.com

^{b)}sajidan@fkip.uns.ac.id

^{c)}harlita@staff.uns.ac.id

Abstract. One competency needed to face 21st-century challenges is critical thinking. In assessing students' critical thinking skills, a proper instrument is needed. This study was aimed to analyze the instrument and find out the students' critical thinking skills. The aspects of critical thinking have been detailed by Facione. They were interpretation, analysis, inference, evaluation, explanation, and self-regulation. The test of critical thinking skills was used as an instrument. This was a quantitative descriptive study conducted on 36 tenth-grade students in Surakarta and Pacitan. Data were analyzed by using Rasch Model with Winstep Software. The reliability of the item was 0.86, which means the instrument has good reliability. Also, validity tests showed the instrument was valid. The validity of the instrument was assessed based on the criteria by Boone: MNSQ, ZSTD, and PTMEASUR CORR. All items were valid. The instrument has good reliability. The results showed the students' critical thinking skills were still low. It is shown from the results of the average logit value of -0.74. The level of difficulty sorted out from the most difficult questions were analysis, evaluation, explanation, self-regulation, interpretation, and inference. Many students cannot give the correct answers to the low difficulty questions as shown by the Wright Map. The results showed that the instrument was valid and reliable and can be used for further research, and there is a need to improve students' critical thinking skills.

Keywords *Critical Thinking, Rasch Model, Ecology*

INTRODUCTION

The Indonesian education system has to prepare a capable and skilled future generation for the nation (Supriyati et al., 2018). Thus, it is necessary to train the 21st-century skills of the students. Critical thinking is one of the basic and important intellectual capital for a human (Zubaidah & Corebima, 2015). Good thinking skills enable humans to organize, adjust, change, or correct their thoughts and thinking process so that they can act more properly (Yustyan et al., 2015). Critical thinking helps to make good decisions and solve complex problems and make breakthroughs in educational achievement (Haseli & Rezaii, 2013).

Critical thinking skills are cognitive processes in systematically analyzing the occurring problems, distinguishing them carefully and thoroughly, and identifying and studying information to plan out the problem-solving strategies (Azizah et al., 2018). (Facione, 2011)), stated that critical thinking was comprised of several aspects or indicators, which are *interpretation, analysis, inference, evaluation, explanation, and self-regulation*. Interpretation is



understanding and expressing the meaning of various experiences, situations, data, events, judgments, rules, procedures, or criteria. The analysis is identifying the expected inferential relationship. The inference is identifying the elements needed to draw reasonable conclusions. Evaluation is assessing the credibility of statements or other representations such as notes or descriptions of perceptions, experiences, or situations. The explanation is stating the evidence, conceptual, or methodological considerations on which one's results are based, and presenting one's reasons in the form of convincing arguments. Self-regulation is the conscious monitoring of personal knowledge by applying skills to analyze and evaluate inferential judgments to themselves.

(York et al., 2018) stated that critical thinking skills were important to be assessed because they are essential abilities and indicators of learning success in achieving the competency standards. Ecology has the potential to be used to train students' critical thinking skills because it relates to the environment. (Saenab et al., 2017). Ecological events or problems can spark students' curiosity and encourage them to make observations or investigations, on which they can acquire new knowledge (Utomo et al., 2016). Ecological events and problems need to be studied through thinking, analyzing, and solving problems by providing proper solutions. Analyzing, solving problems, and providing solutions train critical thinking skills (Harjo et al., 2019).

This research was conducted to determine the validity and reliability of the instrument for assessing students' critical thinking and to find out the reality of students' critical thinking skills. The instrument was tested to find out its normality, validity, and reliability. The data were analyzed using the Rasch Model with Winstep. Paying attention to the results of previous studies and preliminary studies that have been carried out, researchers are encouraged to research analyzing instruments of critical thinking skills that can be used by teachers to measure critical thinking skills of high school students.

METHOD

Data from students' answers were analyzed using the Rasch Model. This was quantitative descriptive research. The samples comprised 34 students taken from high schools in Surakarta and Pacitan. The instruments were distributed directly to schools and via Google Forms. Technique sampling is used in this study is purposive random sampling. The instrument was an essay test about critical thinking skills in Ecology. The instruments were developed from the critical thinking aspect described by Facione, (2011) 1) to comprehend and express the meaning (*interpretation*), 2) to identify the intended and actual inferential relationships (*analysis*), 3) to identify and secure elements needed (*inference*), 4) to assess the credibility of statements or other representation (*evaluation*), 5) to state and justify that reasoning in terms of evidential (*explanation*), and 6) self-consciously to monitor one's cognitive activities (*self-regulation*). Before data collection is carried out, a test instrument is necessary to test using validity and reliability with the tool help Microsoft Excel 2013 and Rasch with Winstep Software. Each aspect was shown in Table 1.

Table 1. Facione's Aspects of Critical Thinking is an Indicator in Research

No.	Aspect	Sub-indicator
1.	Interpretation	a. Categorization b. Interpreting information c. Clarifying information
2.	Analysis	a. Considering various ideas b. Identifying arguments c. Analyzing arguments
3.	Inference	a. Understanding facts b. Constructing premise c. Explaining conclusion
4.	Evaluation	a. Judging fact from information b. Assessing information quality
5.	Explanation	a. Concluding result b. Adjusting to fact c. Presenting arguments
6.	Self-regulation	a. Self-controlling b. Self-correction

The data used in this study are the results of student's critical thinking skills test with 6 indicators integrated into ecology material. The instruments were analyzed using Winstep according to the Rasch Model popularized by Georg Rasch, a Danish mathematician. It has a specialty, that is sample-independent. Boone et al. (2014), stated the criteria to find out the suitability of non-conforming items based on the received Outfit Mean Square (MNSQ) value was $0.5 < \text{MNSQ} < 1.5$; Outfit Z-Standard (ZSTD) conformance to the accepted z test value was $-2.0 < \text{ZSTD} < +2.0$; and Point Measure Correlation (Pt Mean Corr) was $0.4 < \text{Pt Mean Corr} < 0.85$. Rasch's analysis simultaneously ranks the questions in a structured manner from the most difficult to the easiest items and shows the test-taker's ability from highest to lowest.

RESULT AND DISCUSSION

The data were analyzed using the Rasch Model with Winstep. It was aimed to analyze the instrument's validity, instrument reliability, and to determine the achievement of every aspect of critical thinking skills.

Critical Thinking Instrument's Validity

The instrument's validation involved gathering the evidence to provide a scientific basis for the interpretation of test scores. Validity is referring the extent to which evidence and theory support the interpretation of test scores according to the test objectives (Sumintono, 2015). Validity describes the extent to which the measuring instrument (test) measures what has to be measured. The results of the validity test with the Rasch model were shown in Figure 1.

TABLE 10.1 Data Awal Critical Thinking.xlsx ZOU456WS.TXT Sep 21 2020 11:21															
INPUT: 36 PERSON 6 ITEM REPORTED: 36 PERSON 6 ITEM 5 CATS MINISTEP 4.3.2															

PERSON: REAL SEP.: 1.40 REL.: .66 ... ITEM: REAL SEP.: 2.49 REL.: .86															
ITEM STATISTICS: MISFIT ORDER															
ENTRY	TOTAL	TOTAL	MODEL	INFIT	OUTFIT	PTMEASUR	AL	EXACT MATCH							
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	ITEM		

4	44	36	.18	.22	1.50	1.79	1	41	1.48	A	.45	.51	47.2	51.5	Ev4
3	44	36	.18	.22	1.47	1.68	1	41	1.48	B	.48	.51	55.6	51.5	Ex3
5	29	36	1.00	.25	1.36	1.26	1	15	.63	C	.72	.54	38.9	55.1	An5
2	78	36	-1.13	.19	.81	-.86	1	74	-1.10	C	.77	.57	33.3	39.1	If2
1	57	36	-.37	.20	.71	-1.27	1	64	-1.53	B	.66	.54	52.8	47.3	In1
6	45	36	.14	.21	.49	-2.39	1	71	-1.16	A	.62	.51	69.4	51.4	SR6

MEAN	49.5	36.0	.00	.21	1.06	.0	1	1.01	.0				49.5	49.3	
P.SD	15.1	.0	.64	.02	.40	1.6		.33	1.3				11.7	5.1	

TABLE 10.3 Data Awal Critical Thinking.xlsx ZOU456WS.TXT Sep 21 2020 11:21															
INPUT: 36 PERSON 6 ITEM REPORTED: 36 PERSON 6 ITEM 5 CATS MINISTEP 4.3.2															

Figure 1. Critical Thinking Instrument's Validity

Figure 1 shows that an entry number column is a question number (items) arranged from the low validity items to the high validity ones. Meanwhile, the outfit column and the *measure* column show the validity criteria.

(Boone et al., 2014) stated the validity criteria are *output means-square* (MNSQ), *outfit z-standard* (ZSTD), and *point measure correlation* (CORR). An item was valid if it fulfills at least two of the three validity criteria. Based on Figure 1, it can be explained that the item questions meet all the criteria for the *means-square* (MNSQ) with the highest value was 1.41 for Q3 and Q4 and the lowest value was 0.64 for Q1. In the *z-standard outfit criteria* (ZSTD), all items meet all the criteria with the highest value was 1.48 for Q3 and Q4 and the lowest value was -1.53 for Q1. In the *point measure correlation* (CORR), all items meet all the criteria with the highest value was 0.77 for Q2, and the lowest value was 0.45 for Q4. Based on those criteria, all items were valid and no item should be replaced. Therefore, it can be stated that the instrument was valid.



Critical Thinking Instrument's Reliability

In addition to being valid, a test must also be reliable, so that the assessments conducted will get consistent results. The results of the reliability test were shown in Figure 2.

SUMMARY OF 36 MEASURED PERSON									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	8.3	6.0	-.74	.54	.97	-.09	1.01	-.04	
SEM	.7	.0	.18	.02	.13	.19	.14	.19	
P.SD	4.0	.0	1.06	.12	.76	1.15	.82	1.12	
S.SD	4.0	.0	1.07	.12	.77	1.16	.83	1.14	
MAX.	20.0	6.0	1.80	1.10	3.82	2.93	4.16	3.04	
MIN.	1.0	6.0	-3.93	.42	.16	-1.87	.19	-1.73	
REAL RMSE	.61	TRUE SD	.86	SEPARATION	1.40	PERSON RELIABILITY	.66		
MODEL RMSE	.56	TRUE SD	.90	SEPARATION	1.61	PERSON RELIABILITY	.72		
S.E. OF PERSON MEAN = .18									
PERSON RAW SCORE TO MEASURE CORRELATION = .97									
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .71 SEM = 2.15									
SUMMARY OF 6 MEASURED ITEM									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	49.5	36.0	.00	.21	1.06	.04	1.01	-.03	
SEM	6.8	.0	.29	.01	.18	.72	.15	.57	
P.SD	15.1	.0	.64	.02	.40	1.62	.33	1.27	
S.SD	16.5	.0	.71	.02	.44	1.77	.36	1.39	
MAX.	78.0	36.0	1.00	.25	1.50	1.79	1.41	1.48	
MIN.	29.0	36.0	-1.13	.19	.49	-2.39	.64	-1.53	
REAL RMSE	.24	TRUE SD	.60	SEPARATION	2.49	ITEM RELIABILITY	.86		
MODEL RMSE	.21	TRUE SD	.61	SEPARATION	2.83	ITEM RELIABILITY	.89		
S.E. OF ITEM MEAN = .29									

Figure 2. Critical Thinking Instrument's Reliability

Figure 2 shows that the critical thinking skills test instrument has a person reliability value of 0.66, which means the students were inconsistent in answering the questions. The item reliability value was 0.86, which means the items have good reliability. The interaction between person and item can be seen through the Alpha Cronbach value of 0.71. This showed the interaction between person and item was good enough.

Students' Critical Thinking Skill

Students' critical thinking skills were analyzed through the Wright Map analysis. It displays the distribution of students' ability on the left and the distribution of the item difficulty level on the right. The results were shown in Figure 3.

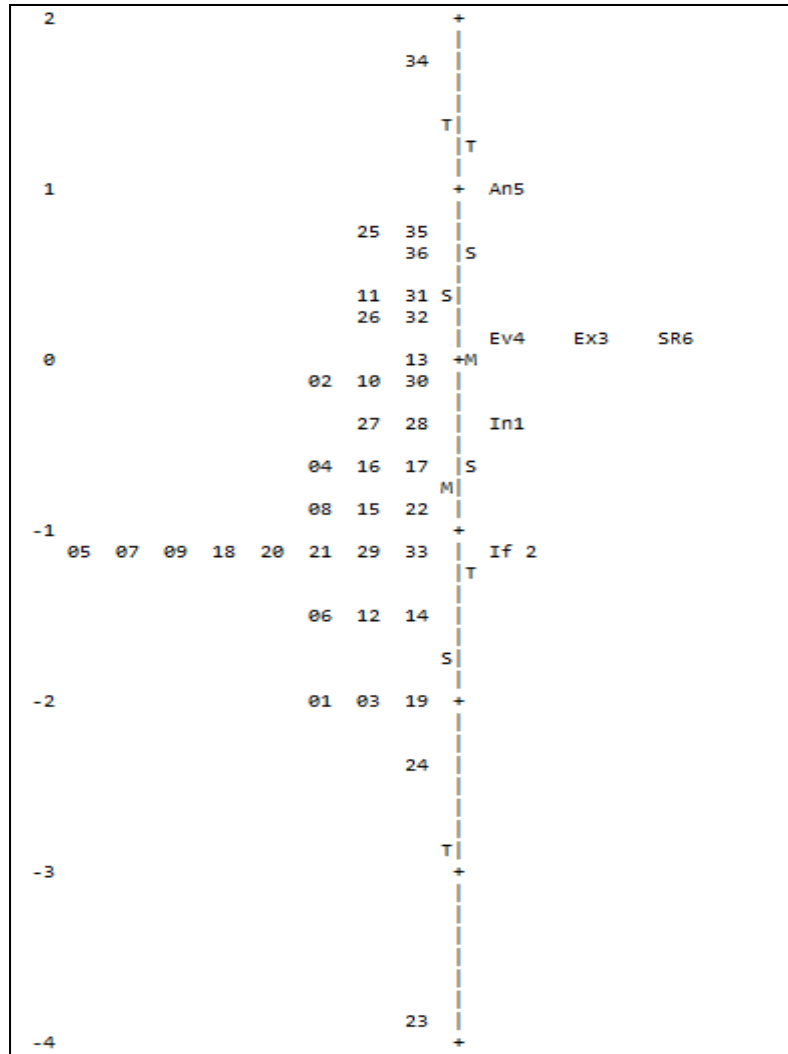


Figure 3. Wright Map

Figure 3 showed the student with the highest ability was S34. This showed that S34 got the maximum score because S34 can answer all the questions. Meanwhile, the student with the lowest ability was S23. Other students with low abilities were S06, S12, S14, S01, S03, S19, and S24. These students could not answer all questions. The item difficulty levels ordered from the most difficult were Q5, Q4, Q3, Q6, Q1, and Q2. Q4, Q3, and Q6 have the same difficulty level. Most students have difficulty answering questions. It can be said, it is necessary to improve the student's critical thinking skills for the progress of their education. The code an5 stands for Q5 analysis, ev4 means Q4 evaluation, ex3 was Q3 explanation, SR6 was Q6 self-regulation, in1 was Q1 interpretation, and if2 was Q2 inference.

a. Analyze's Aspect

Based on the results obtained, the student's critical thinking skills score with the Analyze indicator was the highest. This means that this question is difficult to work on for most students who can answer the analysis questions correctly, namely student number 34. Students can analyze the effects of legume plants planted in rice fields.

b. Evaluation's Aspect

Based on the results of the Rasch analysis, this aspect falls into the moderate question category. With the acquisition that can answer this question, there are 8 students. Students can evaluate questions regarding the impact of fungicide use and the effect of the nutritional cycle by using fungicides and pesticides.



- c. Explanation's Aspect
According to the results of Rasch's analysis, the students' skills in explanation are equivalent to the evaluation aspect. 8 students were able to answer the explanation questions correctly. Students can explain the negative effects due to excessive use of pesticides.
- d. Self-regulation's Aspect
This aspect is quite difficult to work with, but this aspect is equivalent to the aspect evaluation and explanation. Some students can answer this question, namely students number 34, 25, 35, 36, 11, 31, 26, and 32. Students can realize how the carbon cycle works and also the impact of pesticides containing organochlorides on the food chain.
- e. Interpretation's Aspect
Based on the results of Rasch's analysis, this aspect is at the intermediate level. This means that it is neither difficult nor easy. There are 13 out of 34 students who can answer this question. Students can interpret the balance of the ecosystem.
- f. Inference's Aspect
This aspect is at the easiest level. But there are still many students who answer them wrong. Students who answered incorrectly were serial numbers 06, 12, 14, 01, 03, 19, 19, 24, and 23. This means that there are still many students with low critical thinking skills. In this aspect, students are only required to conclude that fish died due to high phosphorus content due to agricultural and industrial waste.

The results showed that the aspect of the analysis is the most difficult item. Items for evaluation, explanation, and self-regulation have the same difficulty level. The item for interpretation was easier than evaluation, explanation, and self-regulation. The easiest item was the inference. It also found eight students who could not answer critical thinking questions properly. Apart from the Wright Map analysis, the student's skills can be seen through the average *logit* value.

PERSON: REAL SEP.: 1.40 REL.: .66 ... ITEM: REAL SEP.: 2.49 REL.: .86

PERSON STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL		INFIT		OUTFIT		PTMEASUR-AL		EXACT OBS%	MATCH EXP%	PERSON
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.				
34	20	6	1.80	.51	.18	-1.87	.19	-1.40	.82	.42	100.0	45.8	34	
25	15	6	.78	.42	1.92	1.66	1.87	1.50	.15	.53	16.7	31.3	25	
35	15	6	.78	.42	2.48	2.34	2.51	2.22	.18	.53	16.7	31.3	35	
36	14	6	.60	.42	.47	-1.18	.46	-1.14	.52	.53	50.0	35.9	36	
11	13	6	.42	.43	.41	-1.35	.45	-1.14	.70	.54	33.3	36.6	11	
31	13	6	.42	.43	.41	-1.35	.45	-1.14	.70	.54	33.3	36.6	31	
26	12	6	.24	.43	.84	-.14	.81	-.18	.80	.54	50.0	37.4	26	
32	12	6	.24	.43	.28	-1.77	.34	-1.45	.59	.54	66.7	37.4	32	
13	11	6	.05	.45	1.99	1.58	2.08	1.63	.63	.53	.0	37.0	13	
2	10	6	-.16	.46	.92	.04	.92	.07	.96	.52	50.0	46.6	2	
10	10	6	-.16	.46	.80	-.19	.80	-.15	.52	.52	66.7	46.6	10	
30	10	6	-.16	.46	.72	-.34	.79	-.18	.70	.52	16.7	46.6	30	
27	9	6	-.38	.48	1.36	.74	1.41	.79	.53	.50	33.3	48.3	27	
28	9	6	-.38	.48	1.36	.74	1.41	.79	.53	.50	33.3	48.3	28	
4	8	6	-.62	.50	3.82	2.93	4.16	3.04	.24	.48	.0	49.5	4	
16	8	6	-.62	.50	.82	-.10	.73	-.25	-.17	.48	33.3	49.5	16	
17	8	6	-.62	.50	.65	-.43	.54	-.66	.10	.48	66.7	49.5	17	
8	7	6	-.88	.53	2.30	1.69	2.63	1.96	.31	.46	16.7	53.8	8	
15	7	6	-.88	.53	1.00	.21	1.00	.22	.12	.46	33.3	53.8	15	
22	7	6	-.88	.53	1.00	.21	1.00	.22	.12	.46	33.3	53.8	22	
5	6	6	-1.19	.57	.67	-.33	.73	-.23	.37	.43	50.0	52.7	5	
7	6	6	-1.19	.57	.16	-1.85	.19	-1.73	.95	.43	83.3	52.7	7	
9	6	6	-1.19	.57	.66	-.36	.71	-.28	.39	.43	50.0	52.7	9	
18	6	6	-1.19	.57	.67	-.33	.73	-.23	.37	.43	50.0	52.7	18	
20	6	6	-1.19	.57	.67	-.33	.73	-.23	.37	.43	50.0	52.7	20	
21	6	6	-1.19	.57	.50	-.71	.47	-.80	.61	.43	50.0	52.7	21	
29	6	6	-1.19	.57	1.97	1.36	2.25	1.65	.21	.43	33.3	52.7	29	
33	6	6	-1.19	.57	.93	.12	.80	-.12	.89	.43	50.0	52.7	33	
6	5	6	-1.54	.61	.21	-1.57	.24	-1.61	.69	.40	100.0	56.3	6	
12	5	6	-1.54	.61	.80	-.10	.82	-.10	.57	.40	66.7	56.3	12	
14	5	6	-1.54	.61	.46	-.79	.48	-.83	.87	.40	66.7	56.3	14	
1	4	6	-1.94	.66	.38	-1.07	.42	-1.05	.65	.37	83.3	54.8	1	
3	4	6	-1.94	.66	.59	-.52	.72	-.32	.20	.37	50.0	54.8	3	
19	4	6	-1.94	.66	.38	-1.07	.42	-1.05	.65	.37	83.3	54.8	19	
24	3	6	-2.43	.73	1.04	.27	.87	-.03	.89	.35	83.3	61.1	24	
23	1	6	-3.93	1.10	1.16	.47	1.27	.58	-.09	.25	83.3	83.7	23	
MEAN	8.3	6.0	-.74	.54	.97	-.1	1.01	.0			49.5	49.3		
P.SD	4.0	.0	1.06	.12	.76	1.1	.82	1.1			25.6	9.6		

Figure 4. Person Measure



Figure 4 showed the logit average value was -0.74 (smaller than 0.0). It indicated that the students were less skilled. This proved the need to improve students' critical thinking skills.

CONCLUSION

The instrument to assess students' critical thinking skills has been produced based on the criteria for critical thinking skills by Facione. The validity of the instrument was assessed based on the criteria by Boone: MNSQ, ZSTD, and PTMEASUR CORR. All items were valid. The instrument has good reliability. The interaction between person and item was good. The level of difficulty sorted out from the most difficult questions were analysis, evaluation, explanation, self-regulation, interpretation, and inference. There was only one student who can answer the questions well and many students unable to give the correct and proper answers. The average student's critical thinking skill was low. So, students' critical thinking skills need to be improved.

Some suggestions that researchers can provide based on the above conclusions are as follows; teachers should develop skills student' critical thinking in a way they often provide training questions and developing learning models which is innovative, especially in terms of critical thinking skills, for students should continue to practice test the ability to think critically by digging up information about relevant arguments and irrelevant arguments to question.

REFERENCES

1. Azizah, M., Sulianto, J., & Cintang, N. (2018). Analisis Keterampilan Berpikir Kritis Siswa Sekolah Dasar Pada Pembelajaran Matematika Kurikulum 2013. 10.
2. Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch Analysis in the Human Sciences. Springer Netherlands. <https://doi.org/10.1007/978-94-007-6857-4>
3. Facione, P. A. (2011). Critical Thinking: What It Is and Why It Counts. *Critical Thinking*, 31.
4. Harjo, B., Harjo, B., Kartowagiran, B., Yogyakarta State University, Indonesia, kartowagiran@uny.ac.id, Mahmudi, A., & Yogyakarta State University, Indonesia, ali_uny73@yahoo.com. (2019). Development of Critical Thinking Skill Instruments on Mathematical Learning High School. *International Journal of Instruction*, 12(4), 149–166. <https://doi.org/10.29333/iji.2019.12410a>
5. Haseli, Z., & Rezaii, F. (2013). The Effect of Teaching Critical Thinking on Educational Achievement and Test Anxiety among Junior High School Students in Saveh. *European Online Journal of Natural and Social Sciences*, 2(2s), 168-175–175.
6. Saenab, S., Pd, S., Pd, M., Yunus, S. R., Pd, S., & Pd, M. (2017). Kementerian Pendidikan Dan Kebudayaan Direktorat Jenderal Guru Dan Tenaga Kependidikan. 23.
7. Sumintono, B. (2015). Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: Implementasi Penilaian Formatif (assessment for learning). 19.
8. Supriyati, E., Setyawati, O. I., Purwanti, D. Y., Sirfa, L., & Prayitno, B. A. (2018). Profil Keterampilan Berpikir Kritis Siswa SMA Swasta di Sragen pada Materi Sistem Reproduksi. 7.
9. Utomo, D. S. W., Si, M., Sutriyono, I., Rizal, D. R., & Si, M. (2016). Pengertian, Ruang Lingkup Ekologi dan Ekosistem. 31.
10. York, T. T., Gibson, C., & Rankin, S. (2018). Defining and Measuring Academic Success. <https://doi.org/10.7275/HZ5X-TX03>
11. Yustyan, S., Widodo, N., & Pantiwati, Y. (2015). Peningkatan Kemampuan Berpikir Kritis Dengan Pembelajaran Berbasis Scientific Approach Siswa Kelas X Sma Panjura Malang. 15.
12. Zubaidah, S., & Corebima, A. (2015). Asesmen Berpikir Kritis Terintegrasi Tes Essay. 15