

Rasch Model Analysis of Critical Thinking Instruments for Elementary School

Kowiyah^{1,3*}, Arita Marini¹, Sihadi Darmo Wihardjo²

{kowiyah_agil@uhamka.ac.id¹, arita250268@yahoo.co.id², sihadwihardjo@gmail.com³}

Elementary Education Department, Universitas Negeri Jakarta Indonesia¹
Department of Environmental Sustainability Education, Universitas Negeri Jakarta, Indonesia²
Elementary Education Department, University of Muhammadiyah Prof.DR.HAMKA Jakarta, Indonesia³

Abstract. One of the 21st century abilities that must be mastered by students is the ability to think critically. To determine the level of successful students in developing critical thinking skills, an instrument is needed that can measure the critical thinking ability. The aimed of this study is to analyze the instrument of critical thinking skills. This research is quantitative method. Data of critical thinking skills was obtained from 269 students on fifth grade from five elementary schools in Jakarta. The data was analyzed by rasch models using winstep version 4.4.3 software. The findings of the pilot study found that the reliability obtained based on the Cronbach Alpha is 0.98, so this value shows instruments used are in very good condition and effectively with a high level of consistency. Reliability of item is 0.99, while the separation of the item is 9.70 and this value can still be used because it shows that the entire item. The respondent was 0.72. While the separation of respondent is 1.62, it indicates in good condition and acceptable. The results showed that all items had a high value of Point Measure correlation which indicated that items could distinguish the ability of respondents. Instruments are valid and reliable can be used for further research

Keywords: Critical Thinking, Instrument, Rasch model.

1. Introduction

Nowadays, education is in the industrial revolution 4.0 faced with the demands of the 21st century. According to the results of research conducted by more than 250 researchers from 60 world institutions members of ATC21S (Assessment & Teaching of 21st Century Skills), there are four skills that must be possessed such as communication, collaboration, critical thinking and problem solving, creativity and innovation. One of 21st century skills is to think critically. Critical thinking is a competency that must be possessed by every individual in globalization era [2], [6], [7]. It is an inseparable part of education because it is a very important cognitive ability [18], so the school keep striving to improve it. Students who have critical thinking will be able to solve problems effectively [3], [10]

Furthermore, the character of the person when facing problem will be seen if he has critical thinking skills. It appears when hespeaks, acts and gives arguments and solutions to a problem. Facione says that there are six main critical thinking skills involved in the process of critical thinking. These skills are interpretation, analysis, evaluation, inference, explanation and self-regulation [4], [11]. In measuring critical thinking skills, an instrument is made based on the aspects of critical thinking. It can be measured through written, oral and observation

tests. Adapted from Starkey, test of critical thinking covers aspects like drawing conclusions, recognizing assumptions, deduction, interpretation, analysis, evaluation of arguments [12].

The competency requirements of industrial revolution 4.0 and 21st century education appears that solving problems and critical thinking is an ability that must be possessed by someone to enter the industrial world as well as a basic foundation for children's thinking skills at the elementary school level. Based on explanation, the researchers conducted research on critical thinking skills in elementary schools by creating instruments to measure students' critical thinking skills in elementary schools. An indicator of the successful learning process is from the score obtained by students. A teacher develops an instrument to get students score. The instrument was arranged based on six aspects of critical thinking skills developed by Focione. It was developed by general events and technical fields faced by students. A good instrument can be trusted and it is measureable. The quality of the instruments is based on the analysis [18].

The analysis was conducted to determine the validity and reliability of the instrument. Validity is related to the accuracy of the assessment tool for what is being assessed so that it really evaluates what should be assessed as the result of the form of the score. There are four kinds of validities, such as content validity, construct validity, predict validity and similarity validity. Content validity relates to the ability of the instrument to measure the content that should be. In this case, the ability of the instrument measures students' critical thinking skills. Construct validity means the ability of an instrument to measure the understanding contained in the material being measured. The ability to predict certain characteristics, behavior or criteria is called the validity of predictions. While the validity of similarity means that the test has similarities with the tests that have been standardized [14]

In this study, the instrument will be analyzed by Rasch modeling using Winstep software. The analysis was carried out so that the resulting instruments had sufficient validity and reliability to measure critical thinking skills. This instrument can be used according to the needs of further research.

2. Methods

In this study, the data was obtained by developing instruments of critical thinking ability assessment. Respondents were 269 students from fifth grade taken from five elementary schools in Jakarta. Instrument was developed from part of one research variable.

The instrument was a questionnaire about critical thinking skills which contained eight questions in the form of open-ended mathematical problem solving questions. Eight questions were developed from six aspects of critical thinking skills delivered by Facione [4] and four aspects were taken which included interpretation, analysis, evaluation, interference. From the four main aspects then it was developed into eight aspects, such as identifying, considering, concluding, communicating data, explaining conclusions, writing results and presenting arguments. In material numbers in elementary school, the instruments of thinking ability were developed from general knowledge and technical knowledge. Then, the respondents determined the correct answer.

The instrument analysis of critical thinking skills used is the Rasch model with the Winstep Software tool version 4.4.3 developed by Linacre (2006)[8]. Analysis of the Rasch model was put forward by Georg Rasch in the 1960s and popularized by Ben Wright. The raw data used is dichotomous data that shows the ability of students. In the analysis of this Rasch model in one model it can be seen the relationship between questions and students [16].

The obtained data was processed using Winsteps software. Rasch model connected student data with questions on the same scale. This scale is obtained from the logarithmic value of the chance that the students are working correctly. By using Rasch model, it can be seen the relationship between the ability level of students (person abilities) and the difficulty level of the problem (difficulty items). Thus, it can be concluded that high-ability students are able to work on easier questions [1].

The Rasch model analysis is able to provide such as the overall information, the quality of the instruments used, the overall response quality of the students, and the interaction between respondents with the items [15]. Person measure shows the average score of all students in working on the problem. Sumintono stated that to measure reliability, it was shown by the value of person reliability and item reliability [16]. Reliability criteria are as follows: <0.67 (weak); 0.67-0.8 (enough); 0.8-0.9 (good); 0.91-0.94 (very good); > 0.94 (excellent).

The analysis of the Rasch model will provide a level of items fit. It will explain whether the item is functioning normally to make measurements. If an obtained question is not fit, this indicates a misconception on the question, so it is useful for the teacher to improve the quality of his teaching [16].

In Rasch model analysis, the quality between item fit and model is abbreviated with item fit. Determination of items fit according to Boone et al in Sumintono (2014) [15] the used criteria are: (1) Value of accepted Outfit mean square (MNSQ) = 0.5 <MNSQ <1.5; (2) The value of the Z-standard Outfit (ZSTD) accepted = -2.0 <ZSTD <+2.0; (3) Value of Measure Correlation (Pr Mean Corr): 0.4 <Pt Measure Corr <0.85.

Analysis of the obtained data using Winsteps software in accordance with the Rasch model if the value of the number of the middle quadrant (mean square) is 1.0 while the standardized value (Z-standarized value) is 0.0. In the analysis, it can also be known that the questions are too easy or too difficult and the respondents are outliers

3. Results and Discussion

In the study, Rasch modeling was carried out with the help of Winsteps Software 4.4.3 to analyze data in order to test the validity and reliability of the instrument. "The Rasch model considers the ability of the respondent to answer each item or question and the level of difficulty of the item itself" (Rasch, 1980). With the analysis of items, compatibility (item fit) can be evaluated whether the items in the instrument can measure what should be measured. Conversely, if the item is not suitable (misfit); then, it is said that the item measures the construct outside the instrument, so the item must be revised or eliminated (Smith, 1992). Based on Wright and Stone (1979).

3.1 Reliability and Separation Items and Respondent

Based on Rasch measurement model approach, the acceptable reliability Cronbach's Alpha is between 0.71-0.99 where it is at the best level (71% - 99%). The findings of the pilot study found that the reliability obtained based on the Cronbach Alpha is 0.98. So this value shows instruments used are in very good condition and effectively with a high level of consistency thus can be used in the actual research.

Person RAW SCORE-TO-MEASURE CORRELATION = .98

CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .70
SEM3.48

Analysis was also performed on the instrument as a whole, namely the reliability and the separation of the respondent. Table 1 shows the reliability and separation respondent where the reliability of the respondent was 0.72, while the separation of respondent is 1.62 when rounded off is equal to 2.0. Based on the reliability of the respondent, “the value of 0.72 indicates are in good condition and acceptable“ [1]. While “ the separation of the item is 1.62 if rounded off is equal to 2.0 and this value can still be used because it shows that the entire item is devide into 2.0 levels of measurement“. According to Linacre (2003) [8], the separation index is better when the value is more than the value of 2.0

Table 1. The Result of Validity and Reliability of Respondent

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	22.8	8.0	.46	.32	1.02	-.01	.99	-.03
SEM	.4	.0	.04	.00	.04	.07	.04	.06
P.SD	6.4	.0	.69	.06	.59	1.13	.72	.98
S.SD	6.4	.0	.70	.06	.59	1.13	.72	.98
MAX.	39.0	8.0	2.60	.77	4.05	3.36	7.30	3.42
MIN.	2.0	8.0	-2.97	.29	.18	-2.51	.17	-2.20
REAL RMSE	.36	TRUE SD	.59	SEPARATION	1.62	Person	RELIABILITY	.72
MODEL RMSE	.32	TRUE SD	.61	SEPARATION	1.89	Person	RELIABILITY	.78
S.E. OF Person MEAN = .04								

Besides, the validity and reliability of respondents in answering questions, questions can also be analyzed from each item to be measured. The following are the results of the reliability of the item question.

Table 2. The Reliability Results of Items

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	766.9	269.0	.00	.06	1.02	-.47	.99	-.60
SEM	74.6	.0	.22	.00	.15	1.79	.13	1.52
P.SD	197.4	.0	.59	.01	.40	4.74	.35	4.03
S.SD	211.1	.0	.63	.01	.43	5.07	.38	4.31
MAX.	1205.0	269.0	.73	.07	1.75	7.41	1.57	5.53
MIN.	515.0	269.0	-1.36	.05	.55	-6.15	.57	-5.36
REAL RMSE	.06	TRUE SD	.59	SEPARATION	9.70	Item	RELIABILITY	.99
MODEL RMSE	.06	TRUE SD	.59	SEPARATION	10.60	Item	RELIABILITY	.99
S.E. OF Item MEAN = .22								

Table 2 shows the reliability of item is 0.99, while the separation of the item is 9.70. This shows that the item are very high reliability and very good. This is because Bond and Fox

“based on the reliability of the item, the value of 0.99 indicates are in very good condition and acceptable“ [1]. While the separation of the item is 9.70 and this value can still be used because it shows that all items

3.2. Items Fit Measure Constructs

Items fit is measuring the constructs that can be seen through “the infit and outfit Mean Square (MNSQ)“.According to Bond and Fox (2007)[1], the outfit and infit MNSQ should be in the range of 0.5 to 1.5 to ensure the items are suitable for measuring the constructs. But the outfit index MNSQ noteworthy in advance compared infit MNSQ for determining congruity of items that measure a construct or latent variable. If the infit or outfit MNSQ value more than 1.5 logit, then it gives meaning confusing item.“ If the MNSQ value is less than 0.5 logit, it shows that the item is too easily anticipated by the respondents“[8], [9]. Point measure correlation should be in the range of 0.4 to 0.85 does not indicate a problem distribution [16]. Besides, the outfit and infit ZSTD value should also be within -2 to +2 [1]. However, if the infit and outfit MNSQ beaccepted, the ZSTD index can be ignored [1], [8]. Applying the Rasch Model: Fundamental Measurement in the Human Sciences.

Table 3. Scale Item Misfit Order

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	Item
3	626	269	.40	.05	1.75	7.41	1.57	5.53	A .56	.55	29.7	33.9	E3
1	793	269	-.04	.05	1.20	2.57	1.37	4.06	B .27	.58	21.9	29.2	E1
2	704	269	.18	.05	1.34	4.03	1.25	2.73	C .60	.56	20.1	32.9	E2
4	1205	269	-1.36	.07	1.23	1.75	1.06	.33	D .62	.58	65.8	65.6	E4
5	910	269	-.34	.05	.78	-3.18	.72	-3.45	d .69	.60	32.7	26.3	E5
6	692	269	.22	.05	.70	-4.37	.70	-3.89	c .65	.56	41.6	32.5	E6
7	690	269	.22	.05	.59	-6.15	.65	-4.76	b .61	.56	35.7	32.4	E7
8	515	269	.73	.06	.55	-5.83	.57	-5.36	a .56	.52	57.2	39.4	E8
MEAN	766.9	269.0	.00	.06	1.02	-.5	.99	-.6			38.1	36.5	
P.SD	197.4	.0	.59	.01	.40	4.7	.35	4.0			15.2	11.5	

Table 3 shows that there are 3 items that are outside the range because they exceed the MNSQ infit and out fit limit values (> 1.5), namely E3 and the ZSTD in fit and outfit limits (> 2), namely E3, E1 and E2). Item E3 has MNSQ infit of 1.75 (> 1.5) and MNSQ 1.57 (> 1.5) out fit and the ZSTD infit and outfit values obtained are also large, namely ZSTD infit of 7.41 (> 2) and ZSTD outfit at 5.53 (> 2). Item E1 has a ZSTD infit value of 2.57 (> 2) and ZSTD outfit is 4.06 (> 2) but the infit and MNSQ outfit values are acceptable, namely MNSQ 1.20 (<1.5) infit and MNSQ outfit 1.37 (<1.5). Furthermore Item E2 has ZSTD infit value of 4.03 (> 2) and ZSTD outfit of 2.73 (> 2) but the infit and MNSQ outfit values can be received, namely MNSQ infit value of 1.34 (<1.5) and MNSQ outfit of 1.25 (<1.5). Thus,there is one item that really is outside the range is E1. Then item E1 needs to be revised or eliminated from the list of items in the research instrument.

3.3 Polarity Item By PTMEA CORR Value

Examination of the Point Measure Correlation (CORR PTMEA) to detect polarity items intended to test the extent to which the construction of constructs to achieve its goal. "If the value contained in the PTMEA CORR is the positive (+), it shows the item measure the constructs to be measured" [1]. Otherwise, if value is negative (-), the item is not developed to measure the constructs to be measured. Thus, it needs to be improved or dropped because the item is not lead to the question (not focus) or difficult to answer by the respondent.

Table 4. Item Polarity Based on Point Measure Correlation

Entry Number	Total Score	Total Count	Measure	Point Measure Corr	Item
8	626	269	0.73	0.56	E8
3	793	269	0.40	0.56	E3
7	704	269	0.22	0.61	E7
6	1205	269	0.22	0.65	E6
2	910	269	0.18	0.60	E2
1	692	269	-0.04	0.27	E1
5	690	269	-0.34	0.69	E5
4	515	269	-1.36	0.62	E4

Based on table 4, it shows that for each item E1 to E8 has a positive Point Measure Correlation value. Thus, there are no items in the instrument that are discarded because "they meet the minimum requirements (PTMEA CORR > 0)" [15]. In addition, the item login value (Measure) which is item E8 of 0.73 shows the most difficult item for the respondent to answer. While E4 items are -1, 36 shows the easiest items to be approved by respondents. The results showed that all items had a high value of Point Measure correlation which indicated that items can distinguish the ability of respondents.

4. Conclusion

From the research that has been done, it can be concluded that there is a typical process of measuring instruments of critical thinking skills in material numbers in fifth grade using the Rasch model. Rasch model analysis provides more comprehensive and in-depth information on the respondent's test, the items simultaneously and accurately. The findings found that the reliability obtained based on the Cronbach Alpha is 0.98, so this value shows instruments that are used in very good condition and with a high level of consistency. The reliability of the item is 0.99 while the item is 9.70 it shows that the entire item. The respondent was 0.72 while the round of off was 1. When rounded off was equal to 2, 0.72 indicated are in good condition and acceptable. The results showed that all items had a high value of Point Measure correlation which indicated that the item could distinguish the ability of the respondents. It can be concluded that the instrument of critical thinking skills could be used for further research.

Reference

- [1] Bond, T.G., & Fox, C.M.(2007). *Applying The Rasch Model: Fundamental Measurement in the Human Sciences, 2nd Edition*. Lawrence Erlbaum Associates, Publishers. Mahwah, New Jersey. London

- [2] Christo Kriel. (2013). Creating a disposition for critical thinking in the mathematics classroom (pp. 67–75). The Biennial Conference of the South African Society for Engineering Education.
- [3] Chukwuyenum, A. N. (2013). Impact of critical thinking on performance in mathematics among senior secondary school students in Lagos State. *IOSR Journal of Research & Method in Education*, 3(5), 18–25. Retrieved from www.iosrjournals.org
- [4] Facione, P. A. (1990). *The California Critical Thinking Skills Test - College Level. Experimental Validation and Content Validity*. California Academic Press. La Cruz Ave. Millbrae
- [5] Fisher, Alec.(2004).*Critical Thinking An Introduction*. Cambridge: University Press.
- [6] Kalelioğlu, F., & Gülbahar, Y. (2014). The effect of instructional techniques on critical thinking and critical thinking dispositions in online discussion. *Educational Technology and Society*, 17(1), 248–258
- [7] Kowiyah. (2016). PENINGKATAN KEMAMPUAN BERPIKIR KRITIS SISWA DALAM PEMECAHAN MASALAH MATEMATIKA MELALUI PENDEKATAN OPEN ENDED Kowiyah1,. *Jurnal Inovasi Pendidikan Dasar*, 2(1), 27–34. Retrieved from <http://jipd.uhamka.ac.id/index.php/jipd/article/view/48>
- [8] Linacre, J. M. (2007). *A User's Guide to WINDTEPS Rasch-Model Computer Programs*. Chicago, Illinois: MESA Press.
- [9] Linacre, J.M. (2010). *User's guide to Winsteps Ministep Rasch-Model Computer Programs*. <http://www.winsteps.com/winman>.
- [10] Peter, E. E. (2012). Critical thinking: Essence for teaching mathematics and mathematics problem solving skills. *African Journal of Mathematics and Computer Science Research*, 5(3), 39–43. <https://doi.org/10.5897/AJMCSR11.161>
- [11] Peter, A., & Facione, P. A. (1989). The California critical thinking skills test -- college level technical report # 4 Interpreting the CCTST , Group Norms , and Sub-Scores
- [12] Stacey, K. (2011). The PISA View of Mathematical Literacy in Indonesia. *Journal on Mathematics Education*, 2(2), 95–126. <https://doi.org/10.22342/jme.2.2.746.95-126>
- [13] Starkey, L. (2009). *Critical thinking skills success teskemampuanberpikir kritis dalam 20 menit*. Yogyakarta: Bookmarks.
- [14] Sudjana. (2005). *Penilaian hasil proses belajarmengajar*. Bandung: PT Remaja Rosdakarya.
- [15] Sumintono, B., & Widhiarso, W. (2013). *Aplikasi Model Rasch Untuk Penelitian Ilmu-ilmu Sosial*. Jakarta: Tim Komunikata Publishing House..
- [16] Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Cimahi: Trim Komunikata Publishing House
- [17] Wibisono, S. (2016). *Aplikasi Model Rasch Untuk Validasi Instrumen Pengukuran Fundamentalisme Agama Bagi Responden Muslim*. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, Vol 5, No. 1, pp. 1-29.
- [18] Zhou, Q., Huang, Q., & Tian, H. (2013). Developing Students' Critical Thinking Skills by Task-Based Learning in Chemistry Experiment Teaching. *Creative Education*, 4(12), 40–45. <https://doi.org/10.4236/ce.2013.412A1006>