5th ICRIEMS Proceedings
Published by Faculty Of Mathematics And Natural Sciences
Yogyakarta State University, ISBN 978-602-74529-3-0

Multiple-Choice With Reason (MCR) To Measure The Critical Thinking Skill On Natural Science : Plant and Animal Movement System

Taufiq Satria Mukti [1,a] and Edi Istiyono [2,b]

[1] *Program Graduate School of Yogyakarta State University*
[2] *Faculty of MIPA of Yogyakarta State University*
Colombo Street No. 1 Karang Malang, Depok, Sleman, Yogyakarta, Indonesia

[a)] Corresponding author: taufiq.satria2016@student.uny.ac.id
[b)]edi_istiyono@uny.ac.id

Abstract. The research conducted is to construct the students' critical thinking skills test instrument of VIII grade natural science subject matter of Plant and Animal Movement System. The developed test instrument is Multiple-Choice With Reason (MCR). Instrument development includes aspects of critical thinking theory adapted to the competencies of Curriculum 2013. *Aiken-V* is used to calculate the instrument feasibility index performed by expert based on construction, grammar, and content with score acquisition .88. The 20 items of Multiple-Choice With Reason (MCR) were tested in 256 VIII grade students at Public Junior School Kendal District. Instrument feasibility test based on empirical data was analyse with *Confirmatory Factor Analysis* (CFA). The results show that Chi-Square value = 225.40, df = 165, P-Value = .0012, RMSEA = .044 so that it can be concluded fit model with 13 items which is feasible to be used.

INTRODUCTION

Assessment is the process of collecting and processing information to measure the achievement of learning outcomes (Mardapi, 2012: 12). Written in the Regulation of the Minister of Education and Culture Year 2016 Number 23 About the Education Assessment Standards that the assessment aims to monitor the development of student learning outcomes and evaluate the learning process. Urgency in performing assessments according (Sudjana, 2013: 3 and Mardapi, 2012: 14) should be based on applicable curriculum operating standards. Thus the instructional objectives in the execution of the assessment goes systematically.

Implementation of education in mengimplemantasikan assessment of critical thinking ability in general is still very terbaikan, which is about 45% (Lane, 2016). The findings of the study (Huber, C.H and Kuncel, N. R. 2016) suggest that the reason for the limitation of critical thinking skills is that it takes longer than the standard has been set. On the other hand, the research done in estimating critical thinking ability generally only focuses on the achievement of value based on test conducted (Anisa, 2017). In addition, not many researchers who studied in depth the grains used and also researchers have not explored the overall ability of students based on the response.

Implementation of a critical thinking skills assessment is often done by using an essay test. It is in accordance with the concept of critical thinking (Ennis, 1986) that in the implementation of the test there needs to be mental involvement, strategy and representation used to solve problems, make decisions, and learn new concepts. However, in fact (McPeck, J. 1981) found that the essay form there are problems that can not be tolerated is the scoring that often leads to the effects of assessment subjectivity.

Critical thinking ability is rarely measured by using multiple-choice test models. This happens because there are too many guesing factors in the implementation (Stephen, 1988) and requires special expertise in making the test items. Nevertheless (Hartini, 2015) and (Akbar et al, 2017) found that critical thinking skills can be measured by multiple-choice tests, noting that the items used are at HOT (High Order Thinking) (Wilson, 2006) and (McPeck, J. 1981) added that the grain must have a high grain difficulty level.

The importance of assessing the ability to think critically in opinion (Travis, 2015) is that critical thinking is an essential ability that can be used as an indicator of learning success in achieving competency standards. In addition, the test is also a form of training in dealing with real-life events in life (Palm, 2008) and also in line with the concept of science learning (Towle, 1989: 16-31) that always put forward critical thinking to be able to

understand each lesson which is very close to real life objects.

Constraints in estimating students' critical thinking skills in the form of essay test and multiple choice need another solution in the form of test with new form. MCR (Multiple-Choice With Reason) is a form of test with options and choosing closed reasons. The form of this test indicates and describes the ability to think critically because it involves the beliefs and knowledge that exist in the student's dilemma to reach the problem and solve complex problems. The critical thinking aspects developed in the MCR test indicators include assumptions, arguments, analysis, evaluation, and conclusions.

In this research article will be discussed about the construction of critical thinking tests based on empirical data. The data analysis is using CFA (Confirmatory Factor Analysis) to see the items that are suitable to use based on existing criteria. The rest of this paper is organized as follow. Section 2 presents a description of the proposed methodology. Section 3 presents the simulation results following discussion. Finally the conclusion and future works are presented in Section 4.

METHODOLOGY

The method used in the construction of the critical thinking assessment test is adopting the Technique of Writing Test Instruments by Mardapi (2016) which includes: 1) Prepare the test specifications. 2) Write a test. 3) Reviewing the test / content validity. 4) Conduct test trials 5) Analyze the test items / construct validity.

The test developed with a Multiple-Choice With Reason (MCR) emphasizing the student to select the answer and selecting the closed reason according to the item context. In the concept of students who really understand the material motion system then students will be able to provide a closed reason that is in the choice. It is also based on the results of the research (...) that it can be concluded that critical thinking ability can be measured by the modified form of multiple choice tests, including multiple choice with reason. The research (Edy, 2010) is also a reference in developing the critical thinking test where in measuring the complex thinking ability can be done with the test type of MCR.

Prior to conducting tests on 256 8th grade students of Public Junior High School in Kendal Regency with criteria of high, medium and low ranking schools based on the value of National Examination of Natural Science subjects in 2017. The items that have been developed must be proved to have a good validity and reliability index. Validity that mean is adjusted to the theory or material Natural Science motion system, construction, materials, and language on the item.Validation is checked by 3 rater (measurement expert, material expert, and practitioner / teacher). Aiken Validity calculation is used to obtain the validity index with the criteria stated either value equal to or greater than .8. While the reliability value with Cronbach Alfa obtained after empirical test item that is stated otherwise if the value equal to or greater than .7.

The scoring done on the item is using the model politomus consisting of 4 categories of scores based on the response of answers and reasons. The scoring of each item is as follows: score of 4 answers and true reason, 3 wrong wrong answer, 2 right answer wrong reason, 1 jawaan and wrong reason. The scoring model of politomus is used because of the different responses (Hambleton, 1991), so that the students' critical thinking ability can be differentiated based on the level of response and the reason in the item (Retnawati, 2014).

Data analysis is done by using application software. Confirmatory factor analysis is used to analyze the validity of the data construct. The fit of all items based on the response model can be seen based on the goodness fit index. The criteria used to assess the goodness of fit are: chi-square normed (χ ^ 2 / df), RMSEA (Root Mean Square Error of Approximation), RMR (Root Mean-square Residual), GFI (Kindness-of-Fit Index) , NFI (Normed Fit Index), Non Fited Fit Index (NNFI), and CFI (Comparative Fit Index). While to determine the fit model of measurement can be done by looking at the index factor Loading and t-Value. The item item is valid and accepted when it has a standard significance value according to Hair, Black, Babin, Ander-son, & Tatham (2010) is "load factor loads of at least ± .3 to .4".

RESULT AND DISCUSSION

The reliability index can be calculated using Cronbach's Alpha statistics before performing construct validity on the instrument. Reliability obtained based on empirical data is .745. The reliability index can be concluded that the instrument has sufficient internal consistency.

The criteria used to look at the suitability of the sleuruh iten model and the suitability of the measurement in the confirmatory factor analysis are: 1. Normed Chi-Square is the ratio between Chi-Square and degrees of freedom. 2. RMSEA is the most informative indicator for the fit model. 3. RMR represents the residual mean by matching the covariance matrix of the data. 4. GFI is a scale of model accuracy that produces a covariance matrix.

5. NFI has a tendency to decrease compatibility in small sample sizes. 6. NNFI is used to correct problems caused by model complexity.

Overall Model Fit

The construct validity analysis can be done after the entire fit model is known. If the overall model complies with the criteria, then the suitability of the measurement model can be done to see the results of construct validity analysis.

The table below shows the overall model fit value after the data is analyzed using the Softwere Application Program.

TABEL 1. the overall model fit

No	GOF Criteria	Result	Level of Fit
1	Normed χ^2	1.27	Good fit
2	RMSEA	.044	Good fit
3	RMR	.093	Marginal fit
4	GFI	.87	Good fit
5	NFI	.37	Marginal fit
6	NNFI	.58	Marginal fit
7	CFI	.64	Marginal fit

Table 1 shows the normed $\chi 2$, that is the ratio between the $\chi 2$ and degree of freedom. Good fit level suggests that the score must be range from 1.00 to 2.00. Because the score is 1.27 so that the normed $\chi 2$ is in the good level. RMSEA scored .044. Because the score is less than the suggested score (RMSEA< .05), so the level of fit is in good fit (Browne & Cudeck, 1993). The result of RMR and CFI is .093 and .64 respectively, with N≤250 the model will in good fit level if Standardized RMR ≤ .09 and CFI> .92 (Hair, Black, Babin, Anderson, & Tatham, 2006) so that those criteria are in marginal fit. The criteria of GFI range from 0 (poor fit) to 1 (perfect fit). The GFI scores .87 so that the fit level is in good criteria. The criteria of GFI is same for NFI and NNFI so that the fit level is in merginal criteria. Considering the condition above, it could say that the overall model is fit to measure the construct validity of the instrument.

Measurement Model Fit

After the overall model is fit, the measurement model fit was conduct. The result of measurement model fit for Critical thinking Skills were described here. Measurement model fit conduct using second order confirmatory factor analysis or 2nd CFA.

The criteria for construct validity are loading factor and *t*-Value. If loading factor (LF) is greater than .3 and *t*-Value more than 1.96 then the item is categorized as valid item. Those criteria refer to assumption that "factor loadings ± .3 to .4 are minimally acceptable." (Hair, Black, Babin, Anderson, & Tatham, 2006).

TABEL 2. First Order CFA

	Sub Aspect	First Order CFA		Interpretation
		LF	t-Val	
CT	A1	.96	2.75	Valid
	A2	1.29	1.97	Valid
	A3	0.88	4.96	Valid
	A4	.73	1.96	Valid
	A5	.83	2.46	Valid

CT : Critical Thinking
A1-A5 : Sub aspect of critical thinking (assumptions, arguments, analyzes, evaluations, and conclusions)

Table 2 shows the summary of construct validity using 1st CFA. It shows that the items of the instrument have loading factor > .3 and t-Value > 1.96 it means the instrument was valid constructively and could measure critical thinking for all Aspect is valid (Hair, Black, Babin, Anderson, & Tatham, 2006).

TABEL 3. Second Order CFA

Sub Aspect	Item	Second Order CFA		Interpretation
		LF	t-Val	
A1	B1	.31	-	Not Valid
	B2	.07	0.66	Not Valid
	B3	.53	2.42	Valid
	B4	.33	2.04	Valid
A2	B5	.16	-	Not Valid
	B6	.51	3.52	Valid
	B7	.31	2.67	Valid
	B8	.32	2.47	Valid
A3	B9	.64	-	Not Valid
	B10	.62	3.21	Valid
	B11	.37	5.99	Valid
	B12	.48	1.98	Valid
A4	B13	.11	-	Not Valid
	B14	.40	2.42	Valid
	B15	.53	2.20	Valid
	B16	.30	2.12	Valid
A5	B17	.29	-	Not Valid
	B18	.82	2.43	Valid
	B19	-.02	.20	Not Valid
	B20	.40	2.12	Valid

A1-A5 : Sub aspect of critical thinking
B1-B20 : Items of critical thinking test

Table 3 shows the summary of construct validity using 2st CFA. It shows that the items of the instrument have loading factor > .3 and t-Value > 1.96 it means the instrument was valid constructively and could measure critical thinking. Valid items are on B3, B4, B6, B7, B8, B10, B11, B12, B14, B15, B16, B18 and B20 (Hair, Black, Babin, Anderson, & Tatham, 2006).

CONCLUSION

Based on analysis result, it can be concluded that conclusion 20 items have factor loading > .3 and t-value > 1.96 so that out of 20 items only 13 items are valid. The 13 items are constructively valid to measure the critical thinking on Natural Science subject especially matter of Plant and Animal Movement System.

The test is particularly useful, both for teachers, who want to monitor students critical thinking toward Natural Science especially matter of Plant and Animal Movement System, and for researchers, who often use different instruments in their studies. For further study, larger samples for empirical data analysis can be performed using other software application programs, and other comparable test models can be developed to obtain a quality instrument for measuring students' criticall thinking skills.

ACKNOWLEDGMENTS

The author thank to the Faculty of Mathematics and Science, Program Graduate School of Yogyakarta State University, and all of contributors for this work.

REFERENCES

1. Akbar, M.N, Firman, H and Rusyati, L. 2017. *Developing Science Virtual Test to Measure Students' Critical Thinking on Living Things and Environmental Sustainability Theme*. IOP Conf. Series: Journal of Physics.
2. Anisa. Aries. 2017. *Meningkatkan Keterampilan Berpikir Kritis Peserta Didik Melalui Pembelajaran IPA Berbasis Potensi Lokal Jepara*. Yogyakarta : UNY/Jurnal Inovasi Pendidikan IPA.
3. Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J.S Long (Eds), *Testing Structural Equatin Models.* Newbury Park, CA: Sage, 136-162 (Electronic Version).
4. Hair, Jr.J.F., Black, W.C., Babin, B.J., Anderson, R.E. & Tatham, R.I. (2006). *Multivarite Data Analysis, 6th edition.* New Jersey: Person Prentice Hall.
5. Hair, J.F., Anderson, R.E., Tatha,, R.L. & Black, W.C. (2010). *Multivariate data analysis* (7th ed). New Jersey: Prentice-Hall.
6. Hambleton, Ronald K. 1991. Fundamentals Of Item Respon Theory. London : New Dehli.
7. Hartini dan Sukardjo. 2015. *Pengembangan Higher Order Thinking Multiple Choice Test Untuk Mengukur Keterampilan Berpikir Kritis Ipa Kelas VII SMP/MTS*. Yogyakarta : UNY/Jurnal Inovasi Pendidikan IPA.
8. Huber, C.H and Kuncel, N. R. 2016. *Does College Teach Critical Thinking? A Meta-Analysis*. Review
a. of Educational Research June, Vol. 86, No. 2, pp. 431 –468.
9. Lane, David and Oswald, Frederick L. 2016. *Do 45% of College Students Lack Critical Thinking Skills? Revisiting a Central Conclusion of Academically Adrif.* Educational Measurement :Issues and Practice Fall, Vol.35, No.3, pp.23–25.
10. McPeck, J. 1981. *Critical Thinking And Education*. New York: St. Martin's.
11. Mardapi, Djemari. 2012. *Pengukuran, Penialian dan Evaluasi Pendidikan*.. Yogyakarta : Nuha Medika. Hlm : 12
12. Moore, Betsy and Stanley, Tood. 2010. *Critical Thinking and formative assesment : increasing the rugor in your classromm*. Eye Education
13. Peraturan Menteri Pendidikan dan Kebudayaan Tahun 2016 Nomor 23 Tentang Standar Penilaian
14. Pendidikan
15. Retnawati, Heri. 2014. Teori Respon Butir dan Penerapannya.Yogyakarta : Parama Publishing Hlm.
16. 32-34.
17. Stephen P. 1988. *Controlling for background belief when developing Multiple-choice CT Test.* Educational Measurement : Issues and Practice, Fall.
18. Sudjana, Nana. 2013. *Penilaian Hasil Proses Belajar Mengajar.* Bandung : PT. Remaja Rosdakarya. Hlm. .3
19. Travis, T. York. 2015. *Defining And Measuring Academic Succes*. The Pennsylvania State University. Practical Assessment, Research & Evaluation, Vol 20, No 5.
20. Towle, A. 1989. *Modern Biology*. Austin: Holt, Rinehart and Winston. Hibbard, K.M. t.t. *Performance Assessment in The Science Classroom*. New York: McGraw-Hill Companies. Hlm 16-31
21. Wilson, Suzanne M and Peterson, Penelope L. 2006. *Theories of Learning and Teaching What Do They Mean for Educators?.* National Education Association. USA : Washington, DC.