



Buddhist critical thinking assessment using Rasch model

Purwo Susongko^{a,*}, Chokchai Yuenyong^b, Arif Zainudin^{c,†}

^a Department of Science Education, Faculty of Teaching Learning and Education, Universitas Pancasakti Tegal, Jawa Tengah 52181, Indonesia

^b Science Education, Faculty of Education, Khon Kaen University, Khon Kaen 40002, Thailand

^c Governmental Studies, Faculty of Social and Politic Science, Universitas Pancasakti Tegal, Jawa Tengah 52181, Indonesia

Article Info

Article history:

Received 19 April 2021

Revised 31 May 2021

Accepted 16 June 2021

Available online 30 April 2022

Keywords:

assessment,

buddha,

critical thinking,

Rasch model

Abstract

There are not many instruments that can be used to measure students 'and adults' critical thinking skills in the real world. Buddhist philosophy teaches 10 indicators to figure out someone's critical thinking ability. This study was intended to; (1) arrange the critical thinking test items based on Buddhist philosophy (Buddhist Critical Thinking Assessment, BCTA); (2) validate the critical thinking test items based on Buddhist philosophy using the Rasch model approach. This study employed the Research and Development approach with ADDIE design. In the development stage, the BCTA test items were tested with 100 respondents from the students of Universitas Pancasakti Tegal, Indonesia consisting of 14 males and 86 females. The BCTA test items were calibrated using the Rasch model with politomos type PCM in three categories. Buddhist Critical Thinking Assessment (BCTA) consists of 10 test items with 10 critical thinking ability indicators based on Buddha in Kalama Sutta. Those 10 items were made in the form of rating scales with three answer choices. The analysis results showed that those BCTA test items have met both content and psychometric validity. The construct validity analysis results showed that all BCTA test items were valid from both content and substantive aspects.

© 2022 Kasetsart University.

Introduction

Critical thinking ability is a very important ability for learning. Through critical thinking, knowledge, understanding, insights and skills can be well obtained in any content. Critical and evaluative analyses are required to learn content. Thus, critical thinking provides a device

to internalize the content and assess the related internalization quality. This enables us to build a system in our mind, to internalize and utilize the system as a reason in solving a problem (Paul & Elder, 2019). In developing higher education in the future, critical thinking is one of three critical thinking skills required in sustainable development (Rieckmann, 2012).

Critical thinking is also essential for citizens to respond to social issues in society (Bermudez, 2015). Critical thinking ability is an essential competency required by the citizens to participate in modern and democratic society. It enables citizens to make a real contribution in society with a full consciousness (Volman,

* Corresponding author.

E-mail address: purwosusongko@upstegal.ac.id (P. Susongko).

† Co-first authors.

E-mail address: arifzainudin@upstegal.ac.id (A. Zainudin).

& ten Dam, 2015). Critical thinking ability can develop the deliberative capacity important for the functions of democratization in healthy conditions (Lim, 2011). With high critical thinking ability, the citizens can easily assess the information and become wise in assessing the government's performance.

Buddhism philosophy is built with critical thinking approach. When religions commonly talk about beliefs in God, Buddhism starts from the discussion about the original causes of human suffering. Buddhism starts from the four proportional statements; (1) Life is suffering; (2) The origin of suffering is craving; (3) The cessation of suffering is *nibbana*; and (4) The noble eightfold path is the way that leads to the cessation of suffering (Batchelor, 2012).

The Buddha also taught to validate all information coming to the students. This is explained in the *Kalama Sutta*. Basically, the *Buddhist teaching* states that the indicators in measuring critical thinking ability can be used in the real world or outside the students' learning class. More complete explanations of those 10 indicators are related to the ability to think critically as proposed by *Buddhist teaching*; (1) Someone should not receive something passed down from generation to generation in spoken form; (2) Someone should not thoughtlessly receive something just because a tradition is passed down from generation to generation; (3) Someone should not blindly receive something since widely spread in public, believed, and agreed by many people; (4) Someone should not receive something as truth just because it is written in a holy book; (5) Someone should not receive something as truth just because it is in accordance with his/her logic; (6) Someone should not receive something as truth just based on hypothesis, prediction, or analysis in mind and hasty in drawing a conclusion; (7) Someone should not receive something as truth just because it makes sense as seen and felt; (8) Someone should not receive something as truth just because based on the previous assumptions; (9) Someone should not receive something as truth just due to credibility, fame, charisma, position, or education from the speaker; and (10) Someone should not receive something as truth just because the speakers are teachers (Bhikkhu, 2012).

At this time, many instruments have been built to assess students' critical thinking ability. However, the weaknesses of those instruments are always made in the specific purposes based on the materials learned by the students and implemented in the form of classroom assessments (Cargas, Williams, & Rosenberg, 2017; Fuad, Zubaidah, Mahanal, & Suarsini, 2017; Phillips, 2019; Widana, Parwata, Parmithi, Jayantika, Sukendra, & Sumandya, 2018). In fact, critical thinking is also required

by students outside the classrooms. The students require critical thinking ability in making better decision, analyzing the information in media carefully, completing work better, not being deceived by any promising product promotion and avoiding some negative elements in our daily life (Butler, 2012). Some researchers have developed large-scale critical thinking assessments, yet still depending on the materials and certain educational levels (Black, 2012; de Bie, Wilhelm, & van der Meij, 2015; Liu, Mao, Frankel, & Xu, 2016; Paul, 2014; Verburgh, François, Elen, & Janssen, 2013).

Most critical thinking assessments are not externally validated to the size reflecting the adults' ways of thinking in real world situations. Most researchers used academic achievements and talent measurements (such as scores, test scores, standard scores), or cognitive ability measurements (such as intelligence test, logical test) to establish its instruments' validity. For example, The Cornell Critical Thinking Test (Ennis, Millman, & Tomko, 2005) compared the scores in assessments using SAT scores and intelligence test scores. The California Measure of Mental Motivation (Insight Assessment Inc., 2011) compared the obtained scores with the measurements of self-efficacy, average scores, and SAT scores. The Watson–Glaser II Critical Thinking Appraisal (NCS Pearson Inc., 2009) also compared the obtained scores with the academic achievement and cognitive competence measurements. The critical thinking assessments aiming at measuring more extensive critical thinking should provide the validity evidence in various life domains. It is difficult to assess critical thinking ability in the real world.

The critical thinking ability assessment instruments can be made in large scale and not dependent on a certain material mastery or education level. Those instruments are expected to be possibly used to measure the critical thinking ability in the real world for students or other adults. Those 10 indicators of critical thinking ability explained in *Kalama Sutta* can be used in developing instruments to measure the critical thinking ability in the real world and can be used outside the class either for students or adults in general.

In developing the instruments, the Rasch model approach can be used to develop the instruments. The Rasch model has recently been considered as the most objective approach in measuring the educational and social assessments. Measurement quality in educational assessments made by the Rasch model will have the same quality such as the measurements made in the physical dimensions in the field of physics (Sumintono & Widhiarso, 2014). The Rasch model is proven appropriate in the critical thinking ability assessments (Jacob, Duffield, & Jacob, 2019; Rachmadtullah, 2020).

To address all these problems, this study aimed at; (1) arranging the Buddhist philosophy-based critical thinking test items (Buddhist Critical Thinking/BCT), (2) validating the BCT test items using the Rasch model approach.

Methodology

This research used a Research and Development Method with the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) approach (Alodwan, & Almosa, 2018). In the analysis stage, the researchers considered the product needs and objectives to develop. In the design stage, the researchers collected, arranged and designed the products to develop. In the development stage, the researchers validated the developed instruments to develop in a small scale. In the implementation stage, the researchers conducted observations by giving assessments to the targeted population with greater sample size. In the evaluation stage, external validity test was conducted by using the appropriate external criteria. This study was limited only to the analysis, design, and development stage.

Analysis

The products of this research were instruments to measure the critical thinking ability to be used by the students in solving the problems in the real world without depending on the mastery of certain courses. The Buddhist philosophy-based critical thinking ability is also known as Buddhist Critical Thinking Assessment (BCTA). BCTA measures the critical thinking ability of adults (more than 18 years old) with the minimum education of Senior High School. BCTA is considered as a potential test instead of a learning result test, whereby it can be tested with anyone without having passed through the learning processes.

Design

BCTA consisted of 10 items representing 10 indicators based on the criteria of someone considered having the critical thinking ability according to Buddha as mentioned in *Kalama Sutta*. Those 10 items were made in the form of rating scales with three answer choices. Each answer choice has score 1, score 2, and score 3. Score 1 shows not critical, score 2 shows adequately critical, while score 3 shows highly critical.

Development

The development stage is a stage to examine the validity of BCTA test items. Thus, BCTA test items were tested with 100 respondents coming from the students of Faculty of Teaching and Education Sciences, Universitas Pancasakti Tegal. The respondents consisted of 79 students learning at the Natural Science Study Program and 21 students from the Mathematics Education Study Program. The respondents consisted of 14 males and 86 females.

The item validity analysis used three validity types consisting of content validity, psychometric validity, and construct validity with Rasch Modeling. The content validity was used by the assessments involving two experts related to the test materials and indicator appropriateness to measure. The assessors were asked to answer whether the test items met the criteria including; (1) item statement correctness, (2) appropriateness of the item choices presented, and (3) correctness of the answer keys. The psychometric validity involved two psychometric experts related to the test construct. The aspect from the assessed test construct covered the material, construction, language, and test narration aspects. Meanwhile, for the construct validity, which referred to the construct validity concept proposed by Messick (1996), the construct validity was divided into six aspects consisting of content, substantive, structural, external, consequential, and generalization (Makransky, Rogers, & Creed, 2015; Ravand, & Firooz, 2016). In the previous studies, the validity of BCTA test items was limited to both the content and substantive validity. This was caused by many obtained data, which were greatly limited and did not cover more extensive population. The calibrated items used the Rasch model with the polytomous types known as the partial credit model (PCM) (formula 2) with 3 categories (1, 2, and 3). Susongko (2016) provided the quantitative criteria related to the construct validity indicators based on the PCM modeling as explained in Table 1.

Table 1 Valid Test Criteria seen from various validity aspects and criteria by implementing the PCM

Construct Validity Aspect	Indicator	Criteria
Content	Item appropriateness	$p > .05$
	Test (itemfit)	$0.5 < \text{MNSQ} < 1.5$ $-2.0 < \text{ZSTD} < 2.0$
	Personal fit statistics	$p > .01$ $0.5 < \text{MNSQ} < 1.5$ $-2.0 < \text{ZSTD} < 2.0$
Substantive		

This research used software utilized for analyzing the PCM modeling using the program of R version 3.5.0 with the package of eRm version 0.16-2. This was open *source* software to be easily accessed and developed by the educational assessment research observers.

Results and Discussion

The research results discuss the formulation of BCTA test items, analysis from all validity aspects and end with a discussion of research results.

Results

Based on the previously arranged test indicators, the arrangement of BCTA test items is presented in the following **Table 2**. Content validity related to the Buddhist philosophy-based critical thinking ability instruments can be implemented because the domains to measure can be clearly figured out and even should pass through the experts' assessments. The experts involved in this research were two people consisting of an expert in science education & Buddhist culture and an expert in Buddhism. From the results obtained from both experts, it can be stated that the BCTA instruments have feasibility or are valid from the content aspect or in accordance with the measurement purposes. From the results assessments made by both psychometric experts, it can be concluded that the BCTA instruments have been made feasible or valid from the psychometric aspect and can be empirically followed with the trials.

Based on the explanation in **Table 1** related to the construct validity criteria on the content aspect, some data resulted from the analysis with the Rasch modeling for the polytomous data, that is, Partial Credit Model (PCM). **Table 3** contains the analysis results on the appropriateness of test items to the model (Item Fit). *Item fit* basically explains whether or not a test item has the function to make the normal measurements.

The test items are quantitatively considered fit or working well if the Outfit MSQ score is between 0.5 to 1.5, the score of outfit *t* is between -2 to 2.0, and the acceptance opportunity of *Ho* (model fitness) is more than 0.05 ($p > .05$). Outfit is the outlier-sensitive fit, a response pattern sensitivity measurement to the test items with certain difficulty levels of the respondents (students) or vice versa. Outfit *t* is the *t*-test for the data appropriateness hypothesis with the model. The value of Outfit MSQ is calculated from the value of chi square divided with the degree of freedom (df). The analysis

results showed that **Table 3** presents all BCTA test items which are appropriate with the used PCM model.

These outfit values illustrated the deviations of responses made by the testees from the ideal model. With the outfit values which are more than the normal limits, it can be stated that those items have quite significant deviations for the Rasch/PCM model. The deviation in this case is that some students who had lower ability than the items' difficulty level could correctly answer those items or other students who had higher ability than the items' difficulty level could not correctly answer those items. The inappropriateness of responses with the model can be caused by many factors, such as carelessness, misconception, guessing success (Sumintono & Widhiarso, 2014). However, the BCTA test items were proven appropriate with the model used.

To see the quality of construct validity from the substantive aspects, the fit test was used to measure the testees' ability to the model. This test basically examined the respondents' consistency or different response patterns to the test items based on the difficulty level. Different response patterns are the inappropriate responses given based on the ability compared to the ideal model. A testee having the ability (θ) of 1.5 should be able to answer all test items with the difficulty level of below 1.5. However, there were some students who were not consistent or resulted in *aberrant responses* in the field. The number of students with *aberrant responses* became the measurement of construct validity with substantive type.

These deviating responses can result from carelessness, *cheating*, or even misconception. Someone's response test whether experiencing deviation or not is called *person fit*. The testees' response acceptance criteria considered deviating or not are the same with the *item fit's* criteria. The testees' qualitative responses considered fit or not experiencing deviation are if the value of MSQ Outfit is between 0.5 up to 1.5, while the value of outfit *t* is between -2 up to 2.0, and the acceptance opportunity of *Ho* (kecocokan model) is bigger than 0.01 ($p > .01$). Of 100 testees, 4 students deviated from the model. It showed that those 4 students did not meet 2 of 3 criteria of person fit (*p* value and MSQ outfit). The list of testees is presented in **Table 4**.

From the above explanations, it can be concluded that 96 percent of testees' responses were normal based on the model or not experiencing deviation, while 4 percent of responses experienced deviation. The percentage of testees with normal responses based on the model can be the bases that the test has met the substantive validity.

Table 2 BCTA test item

No.	Item	Score
1	What is your opinion related to a belief mentioning that 13 is an Unlucky Number? I strongly believe and agree with the opinion. It is necessary to stay cautious with number 13. I do not believe it before having adequate proof.	1 2 3
2	What is your opinion related to a belief on tradition related to good and bad days? I strongly believe and agree with the opinion. It is necessary to stay cautious with those days I do not believe it before having adequate proof.	1 2 3
3	What is your opinion related to the recent news in the social media mentioning that the government has many debts to the other countries? a. I strongly believe and agree with the opinion. b. I will check the truth of the news through social media. c. I do not believe it before having adequate proof.	1 2 3
4	What is your opinion related to the truth of the stories of Holy People mentioned in the Holy Books from various religions? a. I strongly belief b. I will check the truth of those stories based on the scientific data c. I do not believe it before having adequate proof.	1 2 3
5	What is your opinion that every man must marry a woman? I strongly agree b. I agree but I still understand those men who do not get married or have other choices c. I disagree because we have to respect the sexual orientation of each person.	1 2 3
6	All students with outstanding achievements in the class get prizes from their parents. Thomas is one of the students with outstanding achievements in the class, but he looks sad at school. Based on your opinion, why does Thomas look sad? Because Thomas does not get prizes from his parents. Because the prizes given by his parents do not interest Thomas. I can't answer the question before having a direct interview with Thomas.	1 2 3
7	What is your opinion related to two objects with different weight will fall to the earth with different speed? a. I agree. b. I disagree. c. Speed depends on the friction of both objects with the air.	1 2 3
8	Many people assume that girls are generally spoiled. Anna is a girl. What is your opinion related to Anna? a. Anna must be a spoiled girl. b. There is a tendency that Anna has a spoiled character. c. Anna is not necessarily spoiled.	1 2 3
9	Einstein was a world scientist who was wise, brilliant, and can be a good example for the next generations of scientists. What is your opinion of Einstein? a. Einstein's opinion is always correct. b. Einstein's opinion is not always correct. c. Einstein's opinion needs to be proven with the data to know the truth.	1 2 3
10	Whenever joining a lecture, what is your opinion to the related lecturer presenting the materials? a. The lecturer's opinion is always correct. b. The lecturer's opinion is not always correct. c. The lecturer's opinion should be clarified with the other information sources.	1 2 3

Table 3 Analysis results of BCTA test items with the PCM model

Test Item No.	Chisq	df	p	MSQ Outfit	MSQ Infit	Outfit t	Infit t
1	11.882	99	1.000	0.119	0.713	0.47	-0.21
2	83.994	99	.859	0.840	0.902	-0.75	-0.56
3	98.026	99	.509	0.980	0.995	-0.12	0.00
4	89.494	99	.742	0.895	0.889	-0.37	-0.64
5	97.240	99	.531	0.972	0.958	-0.14	-0.28
6	58.135	99	1.000	0.581	0.717	-1.26	-1.99
7	83.466	99	.869	0.835	0.844	-1.03	-0.99
8	78.922	99	.932	0.789	0.852	-1.44	-1.13
9	50.803	99	1.000	0.508	0.763	-1.32	-1.03
10	70.024	99	.988	0.700	0.851	-0.69	-0.48

Table 4 Testees with deviating responses (*aberrant responses*)

Testee	Chisq	df	p	Outfit MSQ	Infit MSQ	Outfit t	Infit t
P05	26.263	9	.002	2.626	2.254	1.76	2.21
P44	33.110	9	.000	3.311	3.089	1.44	2.34
P67	37.455	9	.000	3.745	2.684	1.81	2.34
P69	22.832	9	.007	2.283	2.048	1.63	1.85

The students' deviating responses from the Rasch model show that there was an indication of students who did the test carelessly, used their lucky guess, or even cheating (Sumintono & Widhiarso, 2014). Some studies showed that person fit could be used as the preliminary data related to the students' cheating, carelessness, or lucky guess in completing the test (Meyer & Zhu, 2013; Shu, Henson, & Luecht, 2013). The other studies related to the arrangement of critical thinking ability instruments also showed some inconsistent testees (Harjo, Kartowagiran, & Mahmudi, 2019; Marfu'i, 2019).

Discussion

Based on the content validity and psychometric validity analysis, the BCTA test items can be considered valid and used as items to measure the Buddhist philosophy-based critical thinking ability. However, from the results of construct validity analysis by implementing the Messick validity with the Rasch model approach in the content and substantive aspects, the BCTA instruments have fulfilled the validity criteria. The analysis results are in accordance with the studies previously conducted using the Rasch model to calibrate the test measuring the critical thinking ability (Harjo et al., 2019; Jacob, Duffield, & Jacob, 2019; Japuni & Harun, 2017; Marfu'i, 2019).

BCTA test can be used to assess the critical thinking ability either to the students or other community members. BCTA does not only measure the critical thinking ability based on certain learning material achievement mastery, but also the people's tendency to think critically. The ability to think critically is greatly required by the citizens to support national development (McPeck, 2016; Waller, 2012).

Helpern Critical Thinking Assessment (HCTA) is one test used to measure the critical thinking ability in the real world and widely used already. HCTA consists of 25 daily life scenarios and the testees are asked to analyze and criticize (Butler, 2012; de Bie et al., 2015). The research results on HCTA show that critical thinking is not related to someone's intelligence, yet greatly influences the decisions made by someone (Butler, Pentoney, & Bong, 2017). Some items used in HCTA are

in the form of constructed response test that the testees' scores cannot be completely processed using machine. It is different with BCTA, in which all test items were provided in the forced-choice question format. The question format in BCTA is similarly used in the Cornell Critical Thinking Test (CCTT), namely the forced-choice question format (French, Hand, Nam, Yen, & Vazquez, 2014). The CCTT uses 71 items. A study comparing the validity of HCTA and CCTT proved that HCTA had better content validity than CCTT. Furthermore, the students preferred HCTA to CCTT. However, the use of CCTT required shorter time and was more practical in larger scales (Verburgh et al., 2013).

Based on the comparative studies on HCTA and CCTT, BCTA still appears a more effective critical thinking ability test to use. This is due to many BCTA test items being relatively smaller in number (10 items) and all items were arranged in the forced-choice question format. However, this BCTA is still in preliminary study. More extensive trials or experiments related to the number of involved respondents and respondents' more varied backgrounds are greatly required. To make the quality of BCTA test items perfect, some further studies are necessary to conduct, including: (1) More comprehensive Analysis on Messick validity, (2) Determining the graduation standards and scoring models, (3) Validity test on BCTA scoring criteria, (4) Examining the BCTA test overseas, (5) Adaptation Model of BCTA test for respondents with different languages for communication, (6) Examining the BCTA test with non-student respondents or community groups.

BCTA can be used to measure critical thinking skills in adults. The ability to think for adults is needed to increase productivity in both formal and non-formal education. The application of critical thinking skills in non-formal education in the agricultural sector is proven to increase the economic productivity of farmers (Mariyono, Dewi, Daroini, Latifah, Hakim, & Luther, 2020; Mariyono et al., 2021). In the world of work, this critical thinking ability is very necessary for employees and decision makers to be able to develop innovations in various fields. Therefore, the BCTA can be used for the assessment of employee recruitment and selection of government and private officials. BCTA in the future can

be used as a standard test that can be used widely in accordance with the needs of various parties who require critical thinking skills

Conclusion and Recommendation

The Buddhist Critical Thinking Assessment (BCTA) test items consist of 10 test items based on 10 indicators of critical thinking ability proposed by The Lord Buddha in *Kalama Sutta*. Those 10 test items were made in the form of rating scale with three answer choices. The content experts stated that the BCTA instruments were feasible from the content aspect or based on the measuring purposes. The experts in the psychometric field also stated that the BCTA instruments were feasible from the psychometric aspect. The construct validity analysis results on the content aspect showed that all BCTA test items were appropriate with the used PCM model. The construct validity analysis results on the substantive type showed that 96 percent of respondents were normal based on the model or did not experience deviation, while the other 4 percent of respondents experienced deviation. The analysis results showed that the BCTA test items were valid based on construct seen from both content and substantive aspects.

Conflict of Interest

There is no conflict of interest.

References

Alodwan, T., & Almosa, M. (2018). The effect of a computer program based on analysis, design, development, implementation and evaluation (ADDIE) in improving ninth graders' listening and reading comprehension skills in English in Jordan. *English Language Teaching*, 11(4), 43–51. doi: 10.5539/elt.v11n4p43

Batchelor, S. (2012). A secular Buddhism. *Journal of Global Buddhism*, 13, 87–107. doi: 10.5281/zenodo.1306529

Bermudez, A. (2015). Four tools for critical inquiry in history, social studies, and civic education. *Revista de Estudios Sociales*, (52), 102–118. doi: 10.7440/ress52.2015.07

Bhikkhu, T. (2012). *Kalama sutta: To the Kalamas*. Retrieved from <https://www.accesstoinsight.org/tipitaka/an/an03/an03.065.than.html>

Butler, H. A. (2012). Halpern critical thinking assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology*, 26(5), 721–729. doi: 10.1002/acp.2851

Butler, H. A., Pentoney, C., & Bong, M. P. (2017). Predicting real-world outcomes: Critical thinking ability is a better predictor of life decisions than intelligence. *Thinking Skills and Creativity*, 25, 38–46. doi: 10.1016/j.tsc.2017.06.005

Cargas, S., Williams, S., & Rosenberg, M. (2017). An approach to teaching critical thinking across disciplines using performance tasks with a common rubric. *Thinking Skills and Creativity*, 26, 24–37. doi: 10.1016/j.tsc.2017.05.005

de Bie, H., Wilhelm, P., & van der Meij, H. (2015). The Halpern critical thinking assessment: Toward a Dutch appraisal of critical thinking. *Thinking Skills and Creativity*, 17, 33–44. doi: 10.1016/j.tsc.2015.04.001

Ennis, R. H., Millman, J., & Tomko, T. N. (2005). *Cornell critical thinking tests: Administration manual*. California, CA: Critical Thinking Company.

French, B. F., Hand, B., Nam, J., Yen, H. J., & Vazquez, J. A. V. (2014). Detection of differential item functioning in the Cornell Critical Thinking Test across Korean and North American students. *Psychological Test and Assessment Modeling*, 56(3), 275–286. Retrieved from <https://psycnet.apa.org/record/2014-44772-004>

Fuad, N. M., Zubaidah, S., Mahanal, S., & Suarsini, E. (2017). Improving junior high schools' critical thinking skills based on test three different models of learning. *International Journal of Instruction*, 10(1), 101–116. doi: 10.12973/iji.2017.1017a

Harjo, B., Kartowagiran, B., & Mahmudi, A. (2019). Development of critical thinking skill instruments on mathematical learning high school. *International Journal of Instruction*, 12(4), 149–166. doi: 10.29333/iji.2019.12410a

Insight Assessment Inc. (2011). *Critical thinking attribute tests*. Manuals and assessment information. Retrieved from <https://www.insightassessment.com/wp-content/uploads/ia/pdf/whatwhy.pdf>

Jacob, E. R., Duffield, C., & Jacob, A. M. (2019). Validation of data using RASCH analysis in a tool measuring changes in critical thinking in nursing students. *Nurse Education Today*, 76, 196–199. doi: 10.1016/j.nedt.2019.02.012

Japuni, M. N. B. M., & Harun, J. B. (2017). Validity and reliability of digital games-featured instrument towards critical thinking using the Rasch measurement. *IJSST*, 18(2), 11–15. doi: 10.5013/IJSST.a.18.02.07

Lim, L. (2011). Beyond logic and argument analysis: Critical thinking, everyday problems and democratic deliberation in Cambridge International Examinations' Thinking Skills curriculum. *Journal of Curriculum Studies*, 43(6), 783–807. doi: 10.1080/00220272.2011.590231

Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016). Assessing critical thinking in higher education: The HEIghten™ approach and preliminary validity evidence. *Assessment & Evaluation in Higher Education*, 41(5), 677–694. doi: 10.1080/02602938.2016.1168358

Makransky, G., Rogers, M. E., & Creed, P. A. (2015). Analysis of the construct validity and measurement invariance of the career decision self-efficacy scale: A Rasch model approach. *Journal of Career Assessment*, 23(4), 645–660. doi: 10.1177/1069072714553555

Marfu'i, L. N. R. L. (2019). The analysis critical thinking skills of guidance and counseling students: A pilot study using RASCH model analysis. *Konselor*, 8(2), 52–58. doi: 10.24036/0201982105910-0-00

Mariyono, J., Dewi, H. A., Daroini, P. B., Latifah, E., Hakim, A. L., & Luther, G. C. (2020). Farmer field schools for improving economic sustainability performance of Indonesian vegetable production. *International Journal of Productivity and Performance Management*, Vol ahead-of-print No. ahead-of-print. doi: 10.1108/IJPPM-09-2019-0445

Mariyono, J., Waskito, J., Suwandi, S., Tabrani, T., Kuntariningsih, A., Latifah, E., ... Suswati, E. (2021) Farmer field school: Non-formal education to enhance livelihoods of Indonesian farmer communities. *Community Development*, 52(2), 153–168. doi: 10.1080/15575330.2020.1852436

McPeck, J. E. (2016). *Teaching critical thinking: Dialogue and dialectic*. London, UK: Routledge.

Meyer, J. P., & Zhu, S. (2013). Fair and equitable measurement of student learning in MOOCs: An introduction to item response theory, scale linking, and score equating. *Research & Practice in Assessment*, 8, 26–39. Retrieved from <https://eric.ed.gov/?id=EJ1062822>

NCS Pearson Inc. (2009). *Watson-Glaser II critical thinking appraisal: Technical manual and user's guide*. Retrieved from https://talentlens.in/wp-content/uploads/2017/07/WG2_TechMan_v2_r4.pdf

Paul, S. A. (2014). Assessment of critical thinking: A Delphi study. *Nurse Education Today*, 34(11), 1357–1360. doi: 10.1016/j.nedt.2014.03.008

Paul, R., & Elder, L. (2019). *A guide for educators to critical thinking competency standards: Standards, principles, performance indicators, and outcomes with a critical thinking master rubric*. London, UK: Rowman & Littlefield.

Phillips, J. L. (2019). Making assignments count: The quest for critical thinking in undergraduate political theory essays. *Journal of Political Science Education*, 15(2), 142–160. doi: 10.1080/15512169.2018.1443272

Rieckmann, M. (2012). Future-oriented higher education: Which key competencies should be fostered through university teaching and learning? *Futures*, 44(2), 127–135. doi: 10.1016/j.futures.2011.09.005

Rachmadtullah, R. (2020). Critical Thinking Instrument Test (CTIT): Developing and analyzing Sundanese students' critical thinking skills on physics concepts using Rasch analysis. *International Journal of Psychosocial Rehabilitation*, 24(8), 14426–14443. doi: 10.37200/IJPR/V24I8/PR281423

Ravand, H., & Firoozi, T. (2016). Examining construct validity of the master's UEE using the Rasch model and the six aspects of the Messick's framework. *International Journal of Language Testing*, 6(1), 1–18. Retrieved from https://www.researchgate.net/publication/312093938_Examining_Construct_Validity_of_the_Master's_UEEUsing_the_Rasch_Model_and_the_Six_Aspects_of_the_Messick's_Framework

Shu, Z., Henson, R., & Luecht, R. (2013). Using deterministic, gated item response theory model to detect test cheating due to item compromise. *Psychometrika*, 78(3), 481–497. doi: 10.1007/s11336-012-9311-3

Sumintono, B., & Widhiarso, W. (2014). *Application of the Rasch model for social science research* (Revised edition). Cimahi, Indonesia: Trim Komunita Publishing House.

Susongko, P. (2016). Validation of science achievement test with the Rasch model. *Jurnal Pendidikan IPA Indonesia*, 5(2), 268–277. doi: 10.15294/jpii.v5i2.7690

Verburgh, A., François, S., Elen, J., & Janssen, R. (2013). The assessment of critical thinking critically assessed in higher education: A validation study of the CCTT and the HCTA. *Education Research International*, 2013, 198920. doi: 10.1155/2013/198920

Volman, M., & ten Dam, G. (2015). Critical thinking for educated citizenship. In M. Davies, & R. Barnett (Eds.), *The palgrave handbook of critical thinking in higher education* (pp. 593–603). New York, NY: Palgrave Macmillan.

Waller, N. B. (2012). *Critical thinking: Consider the verdict*. London, UK: Pearson Education, Inc.

Widana, I. W., Parwata, I. M. Y., Parmithi, N. N., Jayantika, I. G. A. T., Sukendra, K., & Sumandya, I. W. (2018). Higher order thinking skills assessment towards critical thinking on mathematics lesson. *International Journal of Social Sciences and Humanities*, 2(1), 24–32. doi: 10.29332/ijssh.v2n1.74