



Development of Critical Thinking Skill Instruments on Mathematical Learning High School

Budi Harjo

Dr Graduate program Yogyakarta Educational Research and Evaluation Yogyakarta 55281, Indonesia, budi.hardjo@student.uny.ac.id

Badrun Kartowagiran

Yogyakarta State University, Indonesia, kartowagiran@uny.ac.id

Ali Mahmudi

Yogyakarta State University, Indonesia, ali_uny73@yahoo.com

The purpose of this study is to develop instruments of critical thinking skills that meet the validity, reliability, and criteria for good items. This research includes development research which consists of the design and implementation stages. The analysis in the study used quantitative analysis with the Confirmatory Factor Analysis (CFA) Technique. The results showed that the value of t value > 1.96 ($\alpha < 0.05$) which means that each significant item measures critical thinking skills. The value of person reliability of 0.75 and item reliability of 0.94 means the consistency of the answers from the good subject. The quality of the items was analyzed using the QUEST program, indicating that the index estimates of the difficulty of the items ranged from -1.65 to 1.55 means that there is no problem that is too difficult or too easy, and is classified as a good item. MNSQ's INFIT value ranges from 0.86 - 1.14 which means that the items match the Rasch model. Outfit value for each item ranged from -1.2 - 1.7 ($t \leq 2.00$) which means that all items were received.

Keywords: development, instruments, critical thinking skills, critical thinking skills instruments, mathematical learning

INTRODUCTION

Critical thinking skills are part of life skills that students must possess, especially in the development of reasoning, communication, and problem-solving faced in students' daily lives. Critical thinking skills in mathematics learning are based on cognitive processes that can be used to solve problems and influence attitudes toward mathematics (Paul, 2007).

Citation: Harjo, B., Kartowagiran, B., & Mahmudi, A. (2019). Development of Critical Thinking Skill Instruments on Mathematical Learning High School. *International Journal of Instruction*, 12(4), .

The competencies and survival skills needed by students in facing the 21st century are emphasized in the following seven skills: (1) critical thinking skills and problem solving, (2) collaboration and leadership, (3) agility and adaptability, (4) initiative and entrepreneurial spirit, (5) able to communicate effectively both orally and in writing, (6) able to access and analyze information, and (7) have curiosity and imagination (Wagner, 2010).

The 21st Century Partnership Learning Framework (P21) (2015) a number of competencies or skills that must be possessed in the 21st century - critical-thinking skills and problem solving, communication and collaboration skills, creative and updating skills, information and communication technology literacy skills, contextual learning skills, as well as information and media literacy skills.

The competencies needed in the 21st century are "The 4Cs", which includes communication, collaboration, critical thinking, and creativity (Kereluik, 2013). Critical thinking skills help students not only learn a set of facts or numbers. Instead, they learn how to find facts and figures for themselves. They ask, and engage in the world around and help others think critically too (Nelson, 2013).

TIMSS measures students' thinking skills towards mathematics subjects. The TIMSS framework divides cognitive dimensions including aspects of knowing, applying, and reasoning. TIMSS mentions that the mathematical problem faced by Indonesian students is the ability to think on cognitive aspects, especially on reasoning (Mullis, 2012).

Critical thinking skills are the ability to combine, mathematical reasoning abilities, generalization skills, ability to prove, and the ability to conduct rational and systematic evaluations. In order to develop critical thinking skills, one must have an attitude and intellectual constellation, which allows the person to get used to critical thinking (Kereluik, 2013).

Critical thinking skills involve activities of observation, reasoning, decision making, analysis, assessment, and conclusion (Cottrell, 2005). Critical thinking skills involve reasoning, analyzing, synthesizing, generalizing and evaluating or conducting evaluation processes (McGregor, 2007). There are seven aspects to measure critical thinking skills, which include seeking the truth, being open-minded, analytical, systematic, confident, having curiosity and maturity (Facione, 1995).

Students' critical thinking skills in several schools in Indonesia are still low. The low level of critical thinking skills of students is marked by symptoms of problems that arise include (1) inaccurate in analyzing a problem; (2) difficult to do high-level questions (C4-C6); (3) passive when doing group work; (4) difficult to connect concepts and problems; (5) difficult expressing their opinions during the discussion. The results showed that aspects of thinking skills in the evaluation aspects were 78%, self-regulation 66%, interpretation 52%, analysis 56%, conclusions 52%, and explanations 42% (Arnita at al., 2019).

The low level of Indonesian students' thinking skills can also be seen from the results of the Program for International Student Assessment (PISA) study which shows that

Indonesia is ranked 69th out of 75 participating countries in 2015 (OECD, 2016). Preliminary studies conducted by researchers throughout July-August 2016 are summarized in Table 1.

Table 1
Summary of Preliminary Study Results

Aspects reviewed	Findings
The instrument developed by the Teacher	Measured students' critical thinking skills
The ability of teachers to develop instruments of critical thinking skills	More than 80% have not been able to develop instruments of critical thinking skills
Level of understanding and critical thinking skills of students	More than 70% of students critical thinking skills are still low
General conclusions	Need to develop instruments that can be used to measure students' critical thinking skills in high school mathematics learning

This study developed the instrument by modifying the Ennis & Weir Critical Thinking Essay Test (EWCTET) model (Ennis & Weir, 1985) and the California Critical Thinking Skill Test (CCTST) model (Facione, 1990). An essay-shaped instrument, used to measure critical thinking skills consisting of five aspects: (1) interpretation, (2) analysis, (3) evaluation, (4) inference, and (5) explanation.

Paying attention to the results of previous studies and preliminary studies that have been carried out, researchers are encouraged to conduct research on developing instruments of critical thinking skills that can be used by teachers to measure critical thinking skills of high school students.

METHOD

Research design

The purpose of this study is to develop instruments of critical thinking skills in high school mathematics learning that is valid, reliable and item quality. This research included development research by modifying the Borg & Gall (1983) and Plomp (2007) models which consisted of the design and implementation phases. The design phase includes the activity of compiling test specifications, writing test items, reviewing and correcting test items, compiling scoring guidelines, and determining completeness criteria. The implementation phase consists of conducting an instrument test, analyzing the item, determining the quality of the test items and describing the test results.

Population and Sample

The study population was high school science program students in Sukoharjo. The study sample was students from nine high schools who were selected with stratified random sampling with the criteria of three schools with good categories, three schools with quite good categories and three schools with poor categories in Sukoharjo Regency. The subject distribution of the trials in this study is presented in Table 2.

Table 2
Distribution of Research Subjects

Subjects	Number of Test Subjects			Amount
	Limited	Readability	Expanded	
Student	30	90	229	349
Teacher	8	6	12	26

Data Analysis

Data analysis in this study applied quantitative analysis. The instruments developed must qualify valid, reliable, and item quality.

Validity

The validity used in this study is content validity and construct validity. Content validation using the Delphi technique involves expert judgment to assess material, construction and language aspects. The instrument used for the validation of the test device consists of 10 items of critical thinking skills in the form of an open-ended. The formula is used to calculate the content validity using the Aiken formulation. (Aiken, 1985). Construct validity provides evidence that the measurement results can actually be interpreted according to the definition, or the instrument actually reveals the measured ability. The validity of the instrument based on empirical data was analyzed using confirmatory factor analysis. The construct validity in this study was tested using first order and second order CFA with the help of LISREL 8.8 software. The first CFA order equation is formulated:

$$\chi_p = \lambda_p \xi + \delta_p \dots\dots\dots (1)$$

where $\chi_1, \chi_2, \dots \chi_p$ is the indicator of the common factor $\lambda_1, \lambda_2 \dots \lambda_p$ is the loading factor $\delta_1, \delta_2, \dots \delta_p$ is the unique factor for each error equation. The relationship between first-order CFA and second-order CFA is formulated in the following equation.

$$\eta = B\eta + \Gamma\xi + \zeta \dots\dots\dots (2)$$

$$X = \Lambda x\eta + \varepsilon \dots\dots\dots (3)$$

Where B is the loading, Γ and Λ coefficient are the first order and second order loading factors, ξ is the latent random vector variable, ε is the residual (Joreskog, 1996).

Reliability

The reliability analysis technique used in this study is multidimensional reliability using Construct Reliability (CR). Construct Reliability (CR) Test according to Hair, et al. (2010) obtained by using the formula:

$$CR = \frac{(\sum_{i=1}^n \lambda_i)^2}{(\sum_{i=1}^n \lambda_i)^2 + (\sum_{i=1}^n \varepsilon)} \dots\dots\dots (4)$$

information:

Harjo, Kartowagiran & Mahmudi

CR = Construct Reliability

λ_i = loading factor of i-indicator standard

ε = standard measurement error

This measure is acceptable if the Construct Reliability (CR) has a value greater than 0.70.

Characteristics of Question Items

This study uses the Rasch model or often called the 1 parameter logistic model (1PL IRT). The Rasch model emphasizes the parameters of the difficulty level of the item (b_i) item. In this study using the theory of item responses for polytomous scoring, with Partial Credit Model (PCM) (Hambleton, 1985). The selection of item criteria is seen from the item parameter values. The level of difficulty of the question (b) has a difficulty index that is in the range of -2 and +2 or is located in the interval 0.3 - 0.7 (Fajrianti et al., 2016). To describe the grain characteristics of critical thinking skills in high school mathematics learning is done with the help of QUEST software.

FINDINGS

Product Development Results

Early Research Results

The initial research phase includes several activities, namely conducting literature studies, preliminary studies, and analyzing problems and needs. In the literature activity, the activities carried out are gathering and documenting supporting research materials relating to the development of instruments, concepts of critical thinking, theories, concepts and postulates relating to the development of instruments of critical thinking skills in high school mathematics learning.

The next activity is to study the results of relevant research so as to obtain research insights related to the theory, procedures, steps, and methods used in development research. Based on the study of the results of relevant research shows that the development of instruments of critical thinking skills in high school mathematics learning is very important. Based on the results of the literature study and preliminary study activities, further analysis of the problem and needs were carried out. The results of identifying problems and needs are presented in Table 3.

Table 3
Summary of Results of Analysis of Problems and Needs

Identification Results	
Problems	Problem Solving Needs
Assessment instruments developed by teachers have not measured critical thinking skills in high school mathematics learning	Teachers need to develop the ability to develop instruments capable of measuring critical thinking skills in high school mathematics learning
Teachers have not been able to develop instruments of critical thinking skills	
The students have low critical thinking skills	Competencies to be developed and measured are emphasized in improving critical thinking skills
Assessment tools designed by teachers cannot be used to measure and improve students' critical thinking skills	

Stage Development Results

In general, the procedure for developing instruments of critical thinking skills is divided into two stages, namely the design and implementation stages. In the initial stages of the design consists of (1) compiling test specifications; (2) writing test items; (3) reviewing and improve the test items; (4) compiling scoring guidelines; (5) determining completeness criteria.

Compiling Test Specifications

The preparation of test specifications includes the act of determining the test objectives, compiling the grid, and selecting the test form. Test specifications serve as a practical guide for the test compiler in planning the content of the subjects being tested, measured behavioral aspects, test forms, and length of the test (Mardapi, 2008). Based on the results of the documentation test obtained three materials measured by the instruments developed, namely: linear programs, sequences and sequences, and systems of equations and not linear equations. The form of the test developed in this study was a test in the form of open-ended.

Writing Test Items

After compiling the test specifications, the next step is to write the test item. Writing test items must consider conformity with the criteria for critical thinking skills tests. In addition, the test items developed should be accompanied by an answer key and guidance on scoring (rubrics). The results obtained at this stage are called draft_1.

Reviewing and Improving Test Items

The next step after writing the test items is to review and improve the test items. The items of the test that have been written (draft 1), the study by involving experts through expert validation. The experts in question are experts consisting of experts in the field of mathematics, measurement, and development of instruments. Based on the results of the analysis, the researcher made a revision. The results of this revision are drawn_2. The revised results based on the input of experts are called draft_3. The results of the revised research instruments, in general, are presented in Table 4.

Table 4
Results of the Research Instrument Revision

Instrument	Number of Previous Items	Corrected	Discarded	Number of Items After Validation
Critical Thinking Skills	10	3	2	8

Compiling Scoring Guidelines

Test casting using reference criteria techniques that are carried out after going through the threshold index calibration process on each item. Interpretation of test scores using item response theory (IRT). The description of the characteristics of the scale to measure the instruments of critical thinking skills is presented in Table 5.

Table 5
Description of Scale Levels of Critical Thinking Skills

Threshold Range	Scale	Category	Characteristic
-1.84 to -.04	0 – 25	Less	Students are less able to do the categorization process, test, identify and analyze ideas, assess statements and arguments, determine facts, estimate alternative answers, describe conclusions and explain the results and procedures correctly
-1.03 to 0.05	26 – 50	Middle	Students can do the categorization process, test and identify ideas, assess statements, estimate alternative answers, but have not been able to analyze ideas, assess arguments, describe conclusions, and explain the results and procedures correctly
0.06 to 0.95	51 – 75	High	Students can do the categorization process, test and identify ideas, assess statements estimating alternative answers, analyze ideas, assess arguments, but not enough to describe conclusions, and explain the results and procedures correctly
0.96 to 1.96	76 – 100	Very high	Students can do the categorization process, test and identify ideas, assess statements estimating alternative answers, analyze ideas, assess arguments, describe conclusions, and explain the results and procedures correctly

For example, based on the results of the expanded trial, information was obtained that in item number 1, there was a test participant with a mean ability of -0.73 which meant that it was in the medium category, indicating that the student could process the categorization, test and identify ideas, assessing statements, estimating alternative answers, but not able enough to analyze ideas, assess arguments, describe conclusions, and explain the correct results and procedures. There were 64 students with a mean ability of 0.58, which means high category, it shows that the student can do the categorization process, test and identify ideas, assess statements estimating alternative answers, analyze ideas, assess arguments, but not enough to describe conclusions, and explain the correct results and procedures. Furthermore, there are 132 students with a mean ability of 1.54 which means that it is a very high category, it shows that the student

can do the categorization process, test and identify ideas, assess statements estimating alternative answers, analyze ideas, assess arguments, draw conclusions, and explain the correct results and procedures.

Determining Completeness Criteria

After drawing up guidelines for test scoring using reference criteria carried out after going through the threshold index calibration process on each item, then the minimum completeness criteria can be determined from each core competency in mathematics learning. This study determined that the minimum threshold of minimum completeness criteria in each core competency based on the threshold range value must be ≥ 0.95 . In other words, the test participant is said to be complete on a core competency if it has a $0.95 \leq$ value threshold range of 96 1.96. For example in item number 1, there are 65 students whose threshold range value ≤ 0.95 and as many as 164 students the threshold range value ≥ 0.95 , thus in item number 1 there are 65 students not completed (28.38%) and 164 completed (71.62%) of all students. In the same way the level of completeness can be determined for all items. The implementation phase consists of: (1) conducting an instrument trial; (2) analyzing items; (3) determining the quality of the test items and describe the test results.

Conducting an Instrument Trial

Before the instrument trial is conducted, the readability, validation and analysis tests are both qualitative and quantitative.

Readability Test

At this stage, the readability of several mathematics teachers was carried out and continued with the initial trial. During the trial, researchers conducted monitoring and interviews with teachers and students to obtain data as a material for revision, reflection and subsequent trial material. Readability test activities involve 8 math teachers who are members of Mathematics Subject Teachers' Consultation in Sukoharjo district. Input from the mathematics teachers on the performance criterion model and research instruments in detail is presented in Table 6.

Table 6
Summary of Readability Test Results

Instrument	Indicator	Mathematics Teacher Input
Critical Thinking Skills	Interpretation	The question of managing critical thinking is seen as having an adequate level of difficulty, not too easy and not too difficult.
	Analysis	
	Evaluation	Questions are considered different from the usual questions for measuring basic competencies. Rubrics need to be prepared that help in measuring the level of critical thinking skills of students.
	Inference	
Explanation		

Validation and Test Analysis

Validation and analysis tests in this study involved experts in the fields of measurement, assessment and mathematics learning to see the extent to which the instruments that had

been developed properly were used to measure thinking skills in high school mathematics learning.

Qualitative Analysis

Aspects that are considered in the qualitative analysis include material aspects, construction and language/readability. Qualitative analysis was carried out through expert validation consisting of three experts in the field of mathematics learning, namely practitioners in the mathematics education study program, Faculty of Mathematics and Natural Sciences, Yogyakarta State University. In general, the results of the qualitative analysis are presented in Table 7.

Table 7
Qualitative Expert Validation

Item	Aspect Material	Category	Construction	Category	Language	Category
1	3.70	Very Worth	3.30	Very Worth	3.70	Very Worth
2	3.30	Very Worth	3.30	Very Worth	3.70	Very Worth
3	2.70	Worth	2.70	Worth	2.70	Worth
4	2.70	Worth	3.00	Worth	3.00	Worth
5	3.00	Worth	2.70	Worth	4.00	Very Worth
6	3.70	Very Worth	3.00	Very Worth	3.30	Very Worth
7	3.70	Very Worth	3.30	Very Worth	3.00	Worth
8	3.00	Worth	3.00	Worth	3.30	Very Worth

The results of the expert validation were confirmed by the product validation criteria submitted by Sultan (2017). Product validation criteria is presented in Table 8.

Table 8
Criteria of Product Validity

Average score	Category
3.26-4.00	Very feasible
2.51-3.25	Feasible
1.76-2.50	Feasible enough
1.00-1.75	Not feasible

Based on Table 7, it can be concluded that based on expert validation, the instrument developed in a feasibility study is used to assess critical thinking skills in high school mathematics learning.

Quantitative Analysis

The aspect measured in the quantitative analysis is the content validity using the Aiken formulation. In general, the value of content validity is presented in Table 9.

Table 9
Content Validity

Item	Aspects of Materials		Aspects of Construction		Aspects of Language	
	Value	Category	Value	Category	Value	Category
1	0.889	High	0.778	Medium	0.889	High
2	0.889	High	0.778	Medium	0.889	High
3	0.556	Medium	0.556	Medium	0.556	Medium
4	0.556	Medium	0.667	Medium	0.667	Medium
5	0.667	Medium	0.556	Medium	1.000	High
6	0.889	High	0.778	Medium	0.778	Medium
7	0.889	High	0.889	High	0.667	Medium
8	0.667	Medium	0.667	Medium	0.778	Medium

Based on Table 9, it can be concluded that the instruments developed in the study are appropriate to be used to assess critical thinking skills in high school mathematics learning. After reading the validity, validation, and analysis of both qualitative and quantitative, an instrument trial was carried out.

Item Analysis

Item analysis were conducted with the aim to determine the suitability of the item model with the Rasch model, the criteria for whether or not the item was accepted, the index of the difficulty of the item, and the construct of the model of measurement of critical thinking skills. Limited trials were conducted on three schools, namely SMAN 1 Sukoharjo, SMAIT Nur Hidayah Sukoharjo, and SMAI Al Azhar 7 Sukoharjo involving 90 students. An analysis is done using QUEST software. The results of this calculation are presented in Table 10.

Table 10
Results of Calculation of Quality Points

Item	Match with Rasch Model			Difficulty Index (Delta value)	
	Infit meansquare	Outfit t	Description	Value	Description
1	0.80	-0.7	Received	1.13	Difficult
2	0.99	0.0	Received	-1.28	Medium
3	1.03	0.7	Received	0.80	Medium
4	1.00	0.0	Received	-0.44	Medium
5	0.82	-1.1	Received	-1.75	Easy
6	1.11	1.7	Received	0.35	Medium
7	1.00	-0.4	Received	0.34	Medium
8	1.15	1.5	Received	-0.18	Medium

Based on Table 10, information can be obtained that from the 8 questions, statistically, it shows that the INFIT value is with a ± 2.0 limit so all items are received. Furthermore, information was obtained that from 8 items there were 1 item or 12.5% of all items that were difficult, 6 items or 75% of all items were medium, and 1 item was difficult or 12.5% of the whole items easy items.

After a limited trial, an expanded trial was conducted at nine schools, namely SMAN 3 Sukoharjo, SMAN 1 Kartasura, Kalam Kudus Christian High School SMAN 1 Nguter,

SMAN 2 Mendungan, As Salam High School, SMAN 1 Mojolaban, Ad Dhuha High School, Pontren Imam Syuhodo High School involving 229 students. The expanded trial was conducted with the aim of knowing the validity and reliability of the instrument. In this study, the First Order and second-order CFA tests were used with the help of LISREL 8.8 software.

First Order CFA Test Instrument Critical Thinking Skills

CFA analysis is calculated using the LISREL 8.8 software. The results of testing the suitability of critical thinking disposition models using first-order CFA are presented in Table 11.

Table 11
Model Match Test Results in Critical Thinking Skills

Value	Limit	Match Index	Description
<i>chi-square</i>	≥ 0.05	$\chi^2 = 3.06$; <i>df</i> = 5; <i>p-value</i> = 0.69138	Fit model
RMSEA	≤ 0.08	0.000	Fit model
NFI	≥ 0.90	1	Fit model

Information about the fit index of each item illustrated from the CFA path first order diagram is presented in Figure 1.

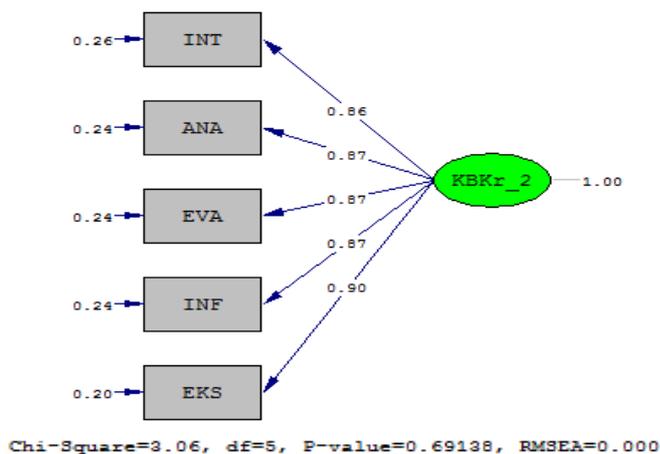


Figure 1
Standard Solution First Order CFA Path Diagram

Information on first order CFA calculations for instruments of critical thinking skills in full is presented in Table 12.

Table 12
CFA First Order Test Results on the Instrument of Critical Thinking Skills

Indicator	Loading Factor	t-value	R ²	Description
Interpretation	0.86	12.54	0.74	Fit
Analysis	0.87	17.87	0.76	Fit
Evaluation	0.87	20.16	0.76	Fit
Inference	0.87	19.31	0.76	Fit
Eksplanan	0.90	18.31	0.81	Fit

Table 12 shows that the fit model measures critical thinking skills. The component that gives the largest contribution to the variable critical thinking skills is the fifth component of the explanation, which is equal to 0.81 while the component that provides the smallest contribution to the variable of critical thinking skills is the first indicator, namely the interpretation of 0.74.

Test for Second Order CFA Instruments Critical Thinking Skills

The results of testing the suitability of the critical thinking skills model through second-order CFA are generally presented in Table 13.

Table 13
Testing Results of Second Order CFA Critical Thinking Skills

Value	Limit	Match Index	Description
chi-square	≥ 0.05	$\chi^2 = 779.66; df = 739; p\text{-value} = 0.14549$	Fit model
RMSEA	≤ 0.08	0.016	Fit model
NFI	≥ 0.90	0.944	Fit model
CFI	≥ 0.90	0.993	Fit model
IFI	≥ 0.90	0.993	Fit model

Determine the Quality of the Test Items and Describe the Test Results

Validity

The empirical validity test was analyzed by the CFA second order test. Based on the calculation of the second order CFA test on each item shows that t value > 1.96 with a confidence level of 95%. This means that each item is statistically significant, or in other words, all items are valid in measuring critical thinking skills in mathematics learning in high school.

Reliability

Reliability calculated in this study is constructed reliability (construct reliability). Construct reliability is calculated by reliability construction test techniques. Based on the calculation results of the construction reliability test, it shows that all components of critical thinking skills are > 0.70. This means that the measurement model of critical thinking skills is in accordance with the data, besides that it is supported by the reliability value of 0.75 people and the reliability of items 0.94 which means the consistency of answers from good subjects.

Quality of test items

The results of the instrument analysis of critical thinking skills involving 229 respondents were tested using the QUEST program. Based on the results of item analysis with the help of the QUEST program, and confirmed by grain quality criteria in accordance with the item response approach, information can be obtained that the index estimation of difficulty items ranges from -1.65 - 1.55. This shows that in terms of the level of difficulty of questions that are not too difficult or too easy, the items in question are classified as good items. MNSQ INFIT values range from 0.86 - 1.14, which means items are in accordance with the Rasch model. The outfit value for each item ranges from -1.2 - 1.7, this indicates the value of Outfit $t \leq 2,00$ which means all items are received. Thus it is concluded that the items about the instrument of critical thinking skills meet the criteria for good quality of goods.

DISCUSSION

Based on the results of the study, information was obtained that the empirical validity test analyzed by the CFA second order test showed $t_value > 1.96$ ($\alpha < 0.05$). This means that each item is statistically significant, or in other words, all items are valid in measuring critical thinking skills in high school mathematics learning. This is relevant to the results of research conducted by Fajrianti et al. (2016) which develops critical thinking tests that can be used in the selection process in the academic and work environment.

Furthermore, the instrument of critical thinking skills in mathematics learning in high school in this study can be used to measure students' critical thinking skills in mathematics learning. The instrument of critical thinking skills developed in this study modified the Ennis & Weir (Ennis & Weir, 1985). EWCTET in the form of open essays used to measure critical thinking skills in high school mathematics learning consisting of five aspects, namely interpretation, analysis, evaluation, inference, and explanation.

The components that influence critical thinking skills based on the results of the CFA second order test analysis include interpretation, analysis, evaluation, inference, and explanation. The component that gives the largest contribution to the variable critical thinking skills is the fifth component of the explanation, which is equal to 0.81 while the component that provides the smallest contribution to the variable of critical thinking skills is the first indicator, namely the interpretation of 0.74. In line with the results of this study, as revealed by Facione et al. (1985) which states that the ability of interpretation and evaluation influences students' critical thinking skills in higher education.

The construct of the instrument of critical thinking skills in high school mathematics learning in this study contains five components, which consist of interpretation, analysis, evaluation, inference, and explanation. This is different from the results of the study by Wechsler (2018) which states that the assessment model of critical thinking skills is explained by five indicators namely deductive reasoning, inductive reasoning, practical reasoning, decision making, problem-solving. This difference is caused by many experts who provide conceptual definitions of critical thinking skills. In line with the results of

research conducted by Rosnawati (2015) who developed a critical thinking evaluation model in learning mathematics in junior high school. There are five aspects of critical thinking skills, namely mathematical reasoning, interpretation, analysis, evaluation, and inference. Researchers, in this case, refers to the EWCTET (Ennis & Weir Critical Thinking Essay Test) model. (Ennis & Weir, 1985)

The model match test is calculated using the QUEST software. Data shows that there are 90 tests analyzed by items as much as 8 on odds 0.5 in accordance with the principle of Likelihood Maximum. From the test results obtained information that the value of person reliability is 0.75 and the value of the item reliability is 0.94. This shows that the consistency of the answers from the subject is good, the quality of the items in the instrument aspects of reliability is good.

Furthermore, by paying attention to the distribution of items based on their suitability with the model, based on the results of item analysis with the help of the QUEST program, and confirmed by grain quality criteria according to the item response approach, information can be obtained that estimates the index of difficulty of items an average of 0.00 and a standard deviation of 0.92. This shows that in terms of the level of difficulty of the questions that are not too difficult or too easy, the items in question are classified as good items. MNSQ's INFIT value ranges from 0.86 - 1.14 which means the item matches the Rasch model. Outfit value for each item ranged from -1.2 - 1.7, this indicates the value of Outfit $t \leq 2,00$ which means all items received. Thus it was concluded that the items fulfilled the quality criteria of the item about the good. This is relevant to the results of the development of instruments of critical thinking skills in this study fulfilling the item criteria of good. The development of critical thinking tests using the item response theory approach (Fajrianti et al., 2016) develops 15 items with fairly good differential power and difficulty levels ranging from -4 to 2.4. The items are of good quality.

Judging from the novelty of the assessment model of mathematics learning in secondary schools, this study serves to measure students' thinking skills in mathematics learning in high school which so far have not been widely developed by teachers, as the efforts of teachers and schools to produce graduates who are able to meet the demands of the 21st century, where critical thinking skills are one of the competencies that must be possessed by students. Critical thinking skills help students in the decision-making process and in solving problems they face in further education and face the challenges of the times because the assessment model is integrated with the learning process and is continuous, so the assessment model is very important to be implemented in mathematics learning in secondary schools.

CONCLUSION

Instruments can actually measure critical thinking skills in a valid manner in measuring critical thinking skills in high school mathematics learning. Each item is statistically significant. The component with the greatest influence on the variable of critical thinking skills is the inference component with a coefficient of determination (R^2) of 95%.

The instrument of critical thinking skills has good reliability, supported by the consistency of answers from good subjects, shown from the results of calculations using the QUEST program the value of person reliability is 0.75 and the value of item reliability is 0.94 and is based on the calculation of measurement of critical thinking skills with construct test reliability which shows that all components of critical thinking skills > 0.70.

The instrument of critical thinking skills meets the criteria of good items, the difficulty of the items about critical thinking skills that are classified as good, the estimated index of difficulty in the questions ranges from -1.65 - 1.55 which means there are no items questions that are too difficult or too easy. MNSQ's INFIT value ranges from 0.86 - 1.14 which means the item matches the Rasch model. The Outfit value for each item ranged from -1.2 - 1.7 or the Outfit value $t \leq 2.00$, so all items were accepted.

REFERENCES

- AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Aiken, L. R. (1985). Three coefficients to analyzing the reliability and validity of rating. *Educational and Psychological Measurement*, 45, 131-142.
- Aizikovitsh, E., & Amit. (2010). *Developing critical thinking skill in mathematics education*. Retrieved from <http://standards.nctm.org/document/chapt2/content.aspx>
- Alcantara, E. C., & Marie, J. P. (2017). Critical thinking and problem-solving skills in mathematics of grade-7 public secondary students. *Asia Pasific Journal of Multidisciplinary Research*, 5/4, 21-27
- Arnita, C. S., Sajidan, Yudi, R., Afandi, & Nanik, M. P. (2019). Improving students' critical thinking skills in cell-metabolism learning using stimulating higher order thinking skills model, *International Journal of Instruction*, 12(2), 327-342
- Asmi, M. N., Waluya, S. B., & Juanedi, I. (2017). Mathematical creative process wallas model in students' problem posing with lesson study approach. *International Journal of Instruction*, 11(2), 527-538
- Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction (4th ed.)*. New York & London: Longman.
- Căprioară, D., & Frunză, V. (2012). Critical Thinking Skills of Students Through Mathematics. *Procedia - Social and Behavioral Sciences* 46(2), 4016-4020.
- Cottrell, S. (2005). *Critical thinking skill: Developing effective analysis and argument*. London: Palgrave Macmillan.
- De Lange, J. (1999). *Framework for classroom assessment in mathematics*. Retrieved from http://www.fi.uu.nl/catch/products/framework/de_lange_frameworkfinal.pdf
- Emir, S. (2009). Education faculty students' critical thinking disposition according to academic achievement. *Procedia Social and Behavioral Sciences*, 1(1), 2466-2469

- Ennis, R. (2011). Critical thinking: Reflection and perspective part I. *Inquiry Critical Thinking Across The Diciplin*, 26(1), 4-18.
- Ennis, R., & Weir, E. (1985). *The Ennis-Weir critical thinking essay test*. CA: Midwest Publications.
- Facione, P. A, Giancarlo, C. A., Facione, N. C., & Gainen, J. (1995). The disposition toward critical thinking. *Journal of General Educationl*. 44(1), 1-25.
- Fajrianthi, Hendriani, W., Septarini, B. G. (2016). Developing critical thinking test utilising item response theory. *Journal of Educational Research and Evaluation*, 20(1), 45-55.
- Glazer, E. (2001). *Using internet primary sources to teach critical thinking skills in mathematics*. London: Greenwood Press.
- Hair, J. F., Black, W., Babin, B.J., & Anderson, R. E. (2010). *Multivariate data analysis*. (7thEd.). NJ: Perason-Prentice Hall, Upper Saddle River.
- Hambleton, R. K., & Swaminantha, H. (1985). *Item response theory*. Boston, MA: Kluwer Njihoff Publishing.
- Halpern, D. F. (2003). *Thought & knowledge: An introduction to critical thinking*. (4th e.d). London: Lawrence Erlbaum Associates Inc.
- Ismail, F., & Kailani, Md. Nor Bin Bakar, Bakry. (2015). developing critical thinking skills of students in mathematics learning. *Journal of Education and Learning*. 9(3), 226-236.
- Joreskog, K. G., & Sorbom, D. (1996). *LISREL 8 structural equation modelling with the simplis command language*. SSI, Software International.
- Kartowagiran, B. (2008). Validasi dimensionalitas perangkat tes ujian akhir nasional SMP mata pelajaran Matematika 2003-2006. *Jurnal Penelitian dan Evaluasi Pendidikan*, 12/2, 177-195.
- Kloppers, M., & Jansen, M. V (2017). Enhancing critical thinking dispositions in the mathematics classroom through a flipped learning approach. *Journal of Communication*, 16(2), 151-160
- Kurniati, S. K., Sabandar, J., & Herman, T. (2014). Mathematical critical thinking ability through contextual teaching and learning approach. *Journal on Mathematic Education*, 6(1), 53-62. doi: dx.doi.org/10.22342/jme.61.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan non tes*. Yogyakarta: Mitra Cendekia Press.
- Mc Gregor, D. (2007). *Developing thinking: developing learning: A guide to thinking skills in educations*. New York: Open University Press.
- Miller, M. D., Linn, R. L., & Grounlund, N. E. (2009). *Measurement and assessment in teaching*. Upper Saddle River: Pearson.

- Mullis, I., Martin, M. O., Rudock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2012). *TIMSS 2011 assessment framework*. Chesnut Hills: TIMSS & PIRLS International Study Center Boston College.
- Moon, J. (2008). *Critical thinking an exploration of theory and practice*. New York: Routledge.
- Nelson, A. C. (2013). Impact of critical thinking on performance in mathematics among senior secondary school students in Lagos state. *IOSR Journal of Research & Method in Education (IOSR-JRME)*. 3(3), 18-25
- NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: The National Council of Teachers of Mathematics, Inc.
- NCTM. (2009). *Guiding principles for mathematics curriculum and assessment*. Retrieved from <http://standards.nctm.org/document/chaptes2/content.aspx?id=23273>.
- Nitko, A. J., & Brookhart, S. M. (2007). *Educational assessment of students*. Boston: Pearson Prentice Hall.
- Husamah, Fatmawati, D., & Setyawan, D. (2017). OIDDE learning model: Improving higher order thinking skills of biology teacher candidates. *International Journal of Instruction*, 11(2), 249-264.
- OECD. (2016). *PISA 2015 result in focus*. PISA: OECD Publishing.
- Partnership for 21st Century Skills (P21). (2015). *Framework for 21st century learning*. Retrieved from P21 Partnership for 21st Century Skills: <http://www.p21.org/our-work/p21-framework>.
- Paul, R., & Elder, L. (2007). *Critical thinking competency standards: standard, principles, performance indicator, and outcomes with critical thinking master rubric*. Berkeley: Foundation for Critical Thinking Press.
- Plomp, T. (2007). *Educational design research: An introduction. An introduction to educational design research*. Proceedings of the seminar conducted at the East China Norma University. Shanghai (PR China), 11, 23-27
- Rosnawati (2015). A Formative evaluation model of critical thinking in teaching mathematics in junior high school. *Journal of Educational Research and Evaluation*, 20(1), 1-10.
- Ruggiero, V. R. (2012). *Beyond feelings: A guide to critical thinking (9th ed.)*. New York: The McGraw-Hill Companies, Inc.
- Saavedra, A., & Opfer, V. (2012). *Teaching and learning 21st century skills: Lessons from the Learning Sciences. A global cities education network report*. New York, Asia Society.

Development of Critical Thinking Skill Instruments on ...

- Shodiq, S., Zamroni, Z., & Kumaidi, K. (2016). Developing an instrument for measuring the faith of the students of Islamic senior high school. *Research and Evaluation in Education*, 2(2), 181-193
- Siegel, H. (2010). Critical thinking. *International Encyclopedia of Education*. 6(1), 1441-1445.
- Smith, C. H., & Cumming, J. J. (2009). *Educational assessment in the 21st century: Connecting theory and practice*. New York: Springer.
- Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, 2(1), 50-80.
- Stein, M. K., Grove, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, 33(2), 455–488.
- Sultan (2017). The development of a critical reading learning model to promote university students' critical awareness. *The New Educational Review*, 76-86.
- Suratno, & Kurniati, D. (2017). The implementation of math-science learning model based on performance assessment to improve students' critical thinking skill in jember coffee plantation area. *Journal of Educational Research and Evaluation*, 21(1), 1-10.
- Wagner, T. (2010). *Overcoming the global achievement gap (online)*. Cambridge, Mass., Harvard University
- Watson, G., & Glazer, E. M. (1994). *Watson-Glazer critical thinking appraisal manual*. San Antonio, TX: The psychologic Corporation: Harcourt Brace Jovanovich.