

# The Assessment of Critical-Thinking-Skill Tests for Accounting Students of Vocational High Schools

Maskhur Dwi Saputra<sup>1</sup>, Soetarno Joyoatmojo<sup>2</sup>, Dewi Kusuma Wardani<sup>3</sup>

## ARTICLE INFO

### Article History:

Received 01.07.2018

Received in revised form

15.08.2018

Accepted

Available online 01.10.2018

## ABSTRACT

This research aimed to develop test-items of critical-thinking skills for Vocational High School (VHS) students of accounting skill program. This research was a research and development with 4-D model (Define, Design, Develop, Disseminate). The sample in this study was VHS students in accounting skill program in Surakarta City, Central Java Province, Indonesia. The sampling technique used cluster random sampling while the data collection technique used questionnaires and test-items. This research developed an essay test consisting of 14 items based on critical-thinking indicators. The results of this study were (1) The assessment of the test-items by experts based on the materials was 91.07%, based on the construction was 71.45%, and based on the language was 80.98%; (2) The results of trials to the students are as follows; (a) There were 8 valid items and 6 invalid items; (b) The level of reliability of the test items was 0.6548 greater than 0.6, so the test items were reliable; (c) There were 5 easy items, 9 medium items, and 0 difficult item; (d) In terms of discriminating power, 6 items were in good category; 2 items were in fair category; and 6 items were in poor category. Based on the results of the expert judgements and trials, it can be concluded that there were 6 feasible items.

© 2018 IJERE. All rights reserved

### Keywords:<sup>1</sup>

Critical-Thinking Skills, Difficulty Level, Discriminating Power, Essay Test, Validation, Reliability.

## INTRODUCTION

The 21<sup>st</sup> century is marked by several skills that everyone must prepare. These skills include: (1) global competition, the need of global awareness and independence; (2) the conditions of global cooperation, the existence of global awareness, cooperation, and the mastery of information and communication technology; (3) the growth of information, the need of awareness to master technology, critical thinking, and problem solving; (4) jobs and careers, the need of critical thinking and problem solving mastery, innovation and refinement, flexibility and adaptability; (5) service-based economic growth namely communication skills, life awareness skills, and careers (Kay, 2008).

Furthermore, there are several required 21<sup>st</sup> century skills: (1) critical thinking (78%); (2) information technology (77%); (3) health and fitness (76%); (4) collaboration (74%); (5) innovation (74%); (6) personal financial responsibility (72%) (Kay, 2008). Based on the data, critical-thinking skills are the most needed skills in the 21<sup>st</sup> century. Therefore, everyone should be able to master critical-thinking skills to be able to compete in the 21<sup>st</sup> century today (Trilling and Fadel, 2009). The mastery of critical-thinking skills can improve the quality of one's life which can be done through education as one of the areas of life to improve human quality (Saeed, Reza, and Momene, 2013). Afsaneh and Tahereh (2015) explained that critical thinking is one of the most fundamental skills in education that needs to be taught to students at every level of education to prepare them to enter the workforce in the 21<sup>st</sup> century.

One level of education to prepare students to enter the workforce is Vocational High Schools (VHS). Students learning at VHS should be able to master the skills in accordance with the skill program they are interested in. The skill programs in VHS are quite a lot, one of which is accounting program. Accounting is interrelated with critical-thinking skills that students who take up accounting must be taught the critical way of thinking in accordance with the science of accounting (Jones, 2010; Lehman, 2013). It was revealed by He, Craig, and Wen (2013) that accounting has some special characteristics such as requiring flexible thinking, analysis, critical thinking, as well as understanding of local and global business scales. In addition, students who take accounting should be able to master the skills of reasoning, processing, and presenting information effectively, productively, collaboratively, and think critically (Directorate of Vocational High School Development, 2017).

<sup>1</sup>Corresponding e-mail: maskhurds.pascauns.fkip@gmail.com, orcid.org/0000-0002-0680-6537  
Faculty of Teacher Training and Education, Sebelas Maret University<sup>1,2,3</sup>

Accounting practices for business are now primarily concerned with analysis, communication, interpersonal and intellectual, so that accounting learning only limited to record, calculate, read, and assess depreciation should be changed into critical-thinking learning (Reinstein and Bayou, 1997). Therefore, students must be given training on critical-thinking skills in order to solve problems properly with rational thinking (Mulyanyo, Gunarhadi, and Indriayu, 2018). Indirectly, critical thinking is also contained in the curriculum of VHS in accounting that accounting students must be able to understand, apply, analyse, and assess factual, conceptual, procedural, and meta-cognitive knowledge (Directorate of Vocational High School Development, 2017). The concept of analysing and assessing contained in the VHS curriculum of accounting is included in critical-thinking skills (Jones, 2010). Based on the description of the characteristics of accounting above, accounting students must be able to master critical-thinking skills.

Critical-thinking skills are included in Higher-Order Thinking Skills (Hashemi et al, 2010; Zubaidah, 2010; Dharmawati, Rahayu, Mahanal, 2017; Purnamawati and Saliruddin, 2017). Higher-order thinking skills are the result of second-level learning (Fitzpatrick and Schulz, 2015), so indirectly critical-thinking skills are also included in the second-level learning outcome. Furthermore, critical-thinking skills are included in the learning outcomes of cognitive domain (Cottrel, 2005, Dharmawati, Rahayu, Mahanal, 2017; Purnamawati and Saliruddin, 2017). Therefore, critical-thinking skills can be measured using test-items (Ennis, 1993; Fisher, 2008). The form of test items that can be used to measure critical-thinking skills is essay test (Syutharidho and Rosida, 2015; Pradana, Parno, and Handayanto, 2016; Dharmawati, Rahayu, Mahanal, 2017). Based on some opinions from the experts above, the measurement of critical-thinking skills in this research used essay test. Test-items were made in accordance with the indicators of critical-thinking skills and applied to the accounting materials in VHS.

### **Critical-Thinking Skills**

Critical thinking is a cognitive activity that includes analysing, assessing, and reconstructing ideas (Paul and Elder, 2002; Fitzpatrick and Schulz, 2015). Activities to properly assess information according to relevant evidence can also be considered as critical thinking (Epstein, 2006; Mason, 2008). In addition, critical thinking is considered as the activities of analysing, synthesizing, evaluating, and concluding (Dwyer, Hogan, and Stewart; 2014). Critical thinking is concerned with active learning to agree or disagree with an idea, evaluate it, and justify it to be a better idea (Florea and Hurjui, 2005). Therefore, critical thinking is an activity of analysing, evaluating or judging, and arguing or creating.

Based on the definition of critical thinking as described above, thinking levels of Bloom's Taxonomy can be used to measure critical-thinking skills because the activities of analysing, evaluating, and creating are included in the cognitive domain of Bloom's Taxonomy. This is supported by some experts who argue that critical thinking can be measured using Bloom's Taxonomy indicators (Pappas, Pierrakos, and Nagel, 2013) such as analysing, synthesizing, and evaluating (Pieterce, Lawrence, and Friedrich-Nel, 2016). Analysing, synthesizing, and evaluating to measure critical thinking were then revised into activities of analysing, evaluating, and creating (DeWealsche, 2015). Analysing is the activity of compiling the parts of the material into a single unit. Evaluating is taking a decision based on certain information while creating is to combine parts to form new ideas (Anderson and Krathwohl, 2001).

### **Research Problems**

Based on the observations in several VHSs in Surakarta that have accounting program, it was found that the test-items made by the teachers were averagely at the levels of remembering (C1) by 28%, understanding (C2) by 48.67%, applying (C3) by 19.33%, analysing (C4) by 4%, evaluating (C5) by 0%, and creating (C6) by 0%. The test-items were not made to direct students to master critical-thinking skills because the thinking level of the questions was still in the domains of remembering, understanding, and applying. The questions to measure critical thinking should be in the domains of analysing, evaluating, and creating (Anderson and Krathwohl, 2001; He, Craig, Wen, 2013; DeWealsche, 2015; Vong and Kaewurai, 2017).

Based on the condition in the field and theoretical review, the researcher developed a test of critical-thinking skills in accounting subject in VHS. Before the test-items were used to measure critical-thinking skills, expert judgements were first performed. The experts gave assessment and feedback on the test-items which include substance, construction, and language (Budiyono, 2015). After the items had been corrected according to the inputs of the experts, they were tested. Trials were conducted by analysing the levels of

validity, reliability, discriminating power, and the difficulty level of each item (Nahadi, Firman, and Kurniadi, 2018).

### **Empirical Study**

This research developed test-items of critical-thinking skills in accounting subject for VHS students. There were several previous researches on the development of critical-thinking skill tests such as the research on students' critical-thinking skills conducted by Aretz, Bolen, and Devereux (2014) by making critical-thinking measurement instruments using CCTDI (California Critical-Thinking Dispositions Inventory). CCTDI test was made using Likert scale (6 points) with 75 test-items. The students answered by choosing one of each given test-items. This CCTDI measuring instrument assumed that students already had ideal critical-thinking skills. The difference with this research is on the form of the instrument used. This research developed critical-thinking instruments by using essay test because not only is critical thinking limited to good or bad, it is also related to the delivery of ideas from students.

Liu, Mao, Frankel, and Xu (2016) conducted a research on the assessment of critical-thinking skills in higher education. The research was applied to the students in higher education, so the widely used test-items covered social science, humanities, natural sciences, and business. The results of this research were the scores of critical-thinking skills. The research was almost the same as that on the development of critical-thinking skills test in this research by developing test-items to measure critical-thinking skills. The difference between the two studies lies in the subjects used. The research on the development of critical-thinking skills test used VHS students of accounting program as its subjects so that the test-items used were related to the accounting phenomena.

A research on the development of critical-thinking test-items in Indonesia was done by Fatimah, Suryani and Yamtinah (2018). They developed critical-thinking skill test-items in History subject for Senior High School students. The difference between these two researches is on the development model used. The previous research used CIPP model (context, input, process, product) while this one used 4-D model (Define, Design, Develop, Disseminate). In addition, Zulfadli (2017) conducted a development research on the test-items based on the revised Bloom's Taxonomy on cell structure and function materials for the eleventh-grade students at Tarakan City Senior High School. The second difference of this research is that the previous research is about Biology test-items. The development of test-items in the previous research used all thinking-levels of the revised Bloom Taxonomy C1 - C6 while this research developed critical-thinking test-items on financial accounting subject with the revised version of Bloom's Taxonomy thinking-levels C4 - C6.

### **Aim of the Study**

This study aimed to develop a feasible test instrument used to measure VHS students' critical-thinking skills in accounting subject. To achieve this objective, there are several formulas that become the discussion in this study; they are: (1) How to develop a test instrument to measure students' critical-thinking skills on accounting subject in VHS ?; (2) Is the development of the test instrument to measure students' critical-thinking skills feasible to use ?.

### **METHOD**

The type of this research was a Research and Development (R&D) using 4-D development model (Define, Design, Develop, Disseminate). The first is defining stage to determine the prerequisites in learning. This stage consists of five steps: preliminary analysis, student analysis, task analysis, conceptual analysis, and determining learning objectives. The second is design stage, which is designing the prototype to be developed. This stage consists of choosing the type of prototype to be developed as needed in the field, designing prototype according to the indicators, and making initial design. The third is development stage conducted to produce a suitable prototype for use in the field. This development stage has several procedures. First, the initial design assessment of the prototype is conducted by the experts. After the prototype design meets the criteria according to the experts, it is tested to the students to determine its feasibility. The last stage is dissemination by using the prototype for wider subjects. At this stage, the prototype can be used to meet students' needs.

This research was conducted at public and private VHSs in Surakarta, Central Java, Indonesia. The sample in this research consisted of three VHSs namely SMKN 1 Surakarta, SMK Batik 1 Surakarta, and SMK Batik 2 Surakarta. The sampling technique used was cluster random sampling while the data collection technique used questionnaires and tests. Questionnaires were used by the experts to assess the developed test-items including substance, construction, and language. The results of the data collection were described following the steps of 4-D development model.

## FINDINGS

This research used 4-D development model (Define, Design, Develop, Disseminate). Below are the results of this research using 4-D model:

### Defining stage

The first stage of the 4-D development model is defining. This stage was done to define the prerequisite of learning which consists of preliminary analysis, student analysis, task analysis, conceptual analysis, and determining the learning objectives.

#### *Preliminary Analysis*

A preliminary analysis was conducted to identify the underlying problems in the field. Based on the observation in the field, there were some problems about the test-items made by the teachers that had not shown critical-thinking skills. The test-items were still at the levels of remembering, understanding, and applying. These levels had not yet led to the level of critical thinking including analysing, evaluating, and creating. Furthermore, this research was conducted in the eleventh-grade of VHS in accounting program in financial accounting subject.

#### *Student Analysis*

This critical-thinking test was tested to 43 students consisting of 4 males and 40 females. The class used was the eleventh grade of accounting department in SMKN 1 Surakarta, Central Java Province, Indonesia. They were 15-17 years old on the average. Piaget in Trianto (2011) stated that students at the age of 15-17 years are able to think in the abstract as well as solve problems through experiments. Therefore, critical thinking can be taught to students through problem-solving activities by using abstract thinking.

#### *Task Analysis*

This research used basic competence 3.7 *Applying merchandise inventory* and 3.8 *Applying inventory methods (FIFO, LIFO, average, and specific identification)* (Directorate of Vocational High School Development, 2017).

#### *Conceptual Analysis*

The concept of inventory subject matter used to generate critical-thinking test-items is illustrated in chart 2 below:

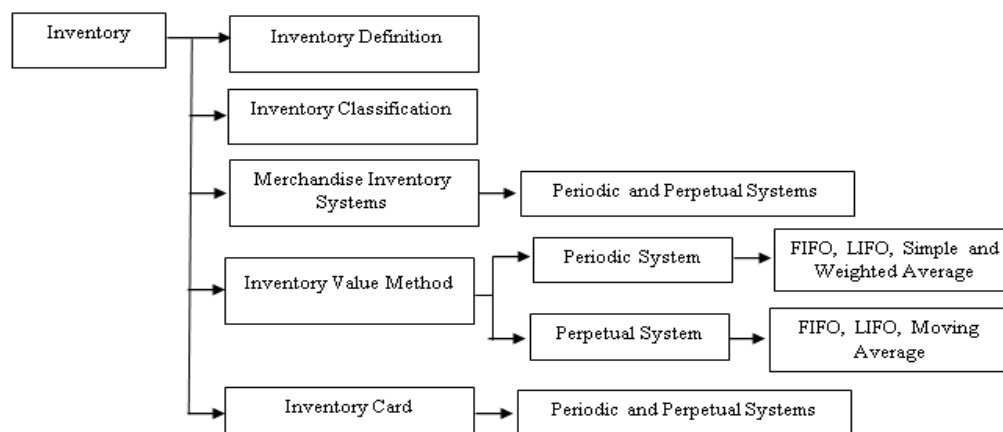


Chart 1. The Conceptual Map of Inventory Subject Matter

### Learning Objectives

Based on the analysis of tasks and concepts as described above, the learning objectives of inventory subject matter are: (1) Students can evaluate the definition of inventory; (2) Students can distinguish inventory classification; (3) Students can distinguish merchandise inventory systems; (4) Students can distinguish methods used to determine inventory values; (5) Students can evaluate the methods used to determine inventory values; (6) Students can create inventory card designs.

### Design Stage

The design stage is the second stage of the 4-D development model. The design of this research was the test-items made based on the indicators of critical thinking (analysing, evaluating, and creating). Below is the instrument outline about critical-thinking test-items:

**Table 1.** The Outline of Critical-Thinking Test Instrument

Indicators of Critical-Thinking Skills	Indicators of Competency Achievement	Test-Items of Critical-Thinking Skills
Analysing (differentiating, organizing, attributing)	Distinguishing the inventory classification and its recording systems.	4, 5, 6, 7, 8
	Distinguishing the methods used to determine inventory values.	9, 10
Evaluating (checking, criticizing, judging)	Evaluating the notion of inventory.	1, 2, 3
	Evaluating the use of methods in determining inventory values.	11, 12
Creating (formulating, planning, producing)	Designing inventory cards to calculate inventory values	13, 14

Sourcer: Anderson and Krathwohl (2001)

Table 1 above is the test instrument outline adjusted to the critical-thinking indicators of the revised version of Bloom's Taxonomy. Each critical-thinking indicator is made into an indicator of achievement of competence that must be mastered by students. Then, the critical-thinking test-items can be made from the indicators of competence achievement with the numbers of items as written in the last column. After the prototype design was described as above, the initial design was made in the form of critical-thinking test-items. The test on this initial design consists of 14 items each of which is representative of every critical-thinking indicator.

### Development Stage

The third stage of the 4-D development model is the development stage, following design stage. At this stage, the assessment of critical-thinking test-items was performed. The assessment was conducted in two ways namely the assessment by experts and trials to students. The first step in the assessment of critical-thinking test-items was done by the experts. It was performed by the expert of evaluation in accounting and the accounting teacher including (1) Materials (substance); (2) Construction; (3) Language.

**Table 2.** The Sub-Indicators of Critical-Thinking Test Assessment

No.	Sub-Indicators
<b>Materials (Substance)</b>	
1	Test-items are according to the Competency Achievement Indicators.
2	The answer keys on each item are correct.
<b>Construction</b>	
3	Test-items are formulated briefly and clearly.
4	Test-items do not lead to multiple interpretations.
5	The answers to the items are not dependent on the answers to the others.

No.	Sub-Indicators
<b>Language</b>	
6	Test-items use good and correct Indonesian.
7	Test-items are made in a communicative language.
8	Test-items do not use local languages.

Sources: The modification from Budiyono (2015)

The accounting evaluation expert and the accounting teacher used the sub-indicators in table 2 above to assess the critical-thinking test-items. The results of the evaluation by the evaluation expert of accounting were: (1) Material aspect: items 1, 2, 3, 4, 5, 9, 10, 11, 12, 13, 14 scored 100%, while items 6, 7, 8 scored 50%; (2) Construction aspect: items 1, 2, 3, 4, 5, 9, 10, 11, 12, 13, 14 scored 66.7%, while items 6, 7, 8 scored 33.3%; (3) Language aspect: items 3, 5, 6, 11, 12 scored 100%, while items 1, 2, 4, 7, 8, 9, 10, 13, 14 scored 66.7%. The overall assessment results of the accounting evaluation expert were: (1) Items 11 and 12 got a total score of 100%; (2) Items 3 and 5 got a total score of 87.5%; (3) Items 1, 2, 6, 9, 10, 13, 14 got a total score of 75%; (4) Item 4 got a total score of 62.5%; (5) Items 7 and 8 got a total score of 50%.

The results of the evaluation by the accounting teacher were: (1) Material aspect: items 1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14 scored 100%, while item 7 and 8 scored 50%; (2) Construction aspect: items 1, 2, 3, 4, 6, 11, 12 scored 100%; items 5, 9, 10, 13, 14 scored 66.7%; and items 7 and 8 scored 33.3%; (3) Language: items 1, 2, 3, 4, 5, 11, 12 scored 100%, while items 6, 7, 8, 9, 10, 13, 14 scored 66.7%. The overall results of the assessment of the critical-thinking test-items done by the accounting teacher were: (1) Items 1, 2, 3, 4, 11, 12 got a total score of 100%; (2) Items 5 and 6 got a total score of 87.5%; (3) Item 9, 10, 13, 14 got a maximum score of 75%; (4) Item 7 and 8 got a total score of 50%.

The results of recapitulation critical-thinking test assessment by the experts were: (1) Material aspect: items 1, 2, 3, 4, 5, 9, 10, 11, 12, 13, 14 scored 100%; item 6 scored 75%; items 7 and 8 scored 50%; (2) Construction aspect: items 11 and 12 scored 100%; items 1, 2, 3, 6 scored 83.4%; items 4, 5, 9, 10, 13, 14 scored 66.7%; items 7 and 8 scored 33.3%; (3) Language aspect: items 3, 5, 11, 12 scored 100%; items 1, 2, 4, 6 scored 83.4%; and items 7, 8, 9, 10, 13, 14 scored 66.7%. The overall total score of items 11 and 12 was 100%; item 3 was 94.5%; items 1, 2, 5 was 88.9%; item 4 was 83.4%; item 6 was 80.6%; items 9, 10, 13, 14 was 77.8%; and items 7 and 8 was 50%. The final score of 14 test-items based on the material indicators was 91.07%; based on the construction indicators was 71.45%; and based on the language indicators was 80.98%.

After the assessment of the experts, the next step was to try out the instrument to the students to determine the feasibility of the test items by using Item Response Theory (IRT) consisting of (1) Validity; (2) Reliability; (3) Discriminating power; (4) Difficulty level (Ayanwale, Adeleke, and Mamadelo, 2018).

### **Validity**

The validation of the critical-thinking test-items was conducted by using Pearson Product Moment formula with assisted application of Microsoft Excel 2013. The result of test-items scoring more than 0.3044 had a valid category. The items scoring above 0.3044 were items 3, 4, 6, 7, 9, 10, 12, 13, while those scoring below 0.3044 were items 1, 2, 5, 8, 11, 14. So, the number of the feasible test-items for this research based on the validation criteria was 8 while the number of infeasible items was 6.

### **Reliability**

The reliability test of the critical-thinking test-items used Cronbach Alpha formula with Cronbach Alpha value criterion greater than 0.6 (Trihendradi, 2012). Based on the calculation of reliability, the overall critical-thinking test-items obtained a total score of 0.6548. Hence, the critical-thinking test-items obtained a total score above Alpha Cronbach of 0.6, so the items were feasible to use based on the criterion of reliability. The reliability criteria are based on the calculation method used, namely the Cronbach Alpha method. The reliability test with this method sets the reliability criteria of above 0.6 (Trihendradi, 2012). In addition, the reliability criteria were adjusted to the students' abilities in Indonesia.

### *Discriminating Power*

The discriminating-power test of critical-thinking test-items used the same formula as the validation test that is Pearson Product Moment, so the result of the discriminating-power test would be the same with the validation test. The feasible test-items for this research based on the discriminating-power test were those scoring  $\geq 0.40$  or  $0.30-0.39$ . The calculation results of the discriminating-power test were: (1) the items scoring  $\geq 0.40$  were items 6, 7, 9, 10, 12, and 13; (2) the items scoring  $0.30-0.39$  were items 3 and 4; (3) the items scoring  $< 0.30$  were items 1, 2, 5, 8, 11, 14. Therefore, the number of feasible test-items for this research based on the discriminating-power criteria was 8 while the number of infeasible items was 6.

### *Difficulty Level*

The difficulty level of critical-thinking test-items was calculated by dividing the average for each score of the item with the maximum score of each item. Test-items for this research based on the difficulty-level criteria scored  $0.31-0.71$ . Based on the calculation results of the difficulty level, the items scoring  $0.31-0.71$  were items 1, 2, 3, 4, 7, 9, 11, 12, 13 while those scoring above  $0.71$  and below  $0.31$  were items 5, 6, 8, 10, 14. Hence, the number of feasible test-items for this research based on difficulty-level criteria was 9 while the number of infeasible items was 5.

The results of the validation test, reliability test, discriminating-power test, and difficulty-level test are summarized in table 3 below:

**Table 3.** *The Feasibility of Critical-Thinking Test-Items*

Item Numbers	Test-Item Feasibility				Feasibility Criteria
	Validity Test	Reliability Test	Discriminating Power Test	Difficulty Power Test	
1	Invalid	Reliable	Poor	Medium	Infeasible
2	Invalid	Reliable	Poor	Medium	Infeasible
3	Valid	Reliable	Fair	Medium	Feasible
4	Valid	Reliable	Fair	Medium	Feasible
5	Invalid	Reliable	Poor	Easy	Infeasible
6	Valid	Reliable	Good	Easy	Infeasible
7	Valid	Reliable	Good	Medium	Feasible
8	Invalid	Reliable	Poor	Easy	Infeasible
9	Valid	Reliable	Good	Medium	Feasible
10	Valid	Reliable	Good	Easy	Infeasible
11	Invalid	Reliable	Poor	Medium	Infeasible
12	Valid	Reliable	Good	Medium	Feasible
13	Valid	Reliable	Good	Medium	Feasible
14	Invalid	Reliable	Poor	Easy	Infeasible

Sources: The processed data summary of the critical-thinking test-item trials

Table 3 above is the criteria of critical-thinking test-items that have been tested. The test-items considered feasible were those which were valid, reliable, had very good and good discriminating power, and had a medium level of difficulty. The test-items that met the four criteria were items 3, 4, 7, 9, 12, and 13. So, the test-items used in this research were 6 items.

### **Dissemination Stage**

The last stage of the 4-D development model is the dissemination. This stage was done by disseminating the test-items to wider schools. Below are the feasible test items to disseminate to measure students' critical-thinking skills.

**Table 4. Feasible Critical-Thinking Test Items**

Indicators of Critical-Thinking Skills	Indicators of Competency Achievement	Test-Items of Critical-Thinking Skills
Analysing (differentiating, organizing, attributing)	Distinguishing the inventory classification and its recording systems.	4, 7
	Distinguishing the methods used to determine inventory values.	9, 10
Evaluating (checking, criticizing, judging)	Evaluating the notion of inventory.	3
	Evaluating the use of methods in determining inventory values.	12
Creating (formulating, planning, producing)	Designing inventory cards to calculate inventory values	13

Sources: Anderson and Krathwohl (2001)

The number of feasible test-items was 6, and they had represented each indicator of critical thinking as well as the indicators of competence achievement. The items were used to measure the critical-thinking skills of VHS students of accounting skill program. The schools used to disseminate the test-items were SMK Batik 1 Surakarta and SMK Batik 2 Surakarta in Central Java Province, Indonesia. The items were disseminated to the eleventh-grade students of accounting in both VHSs. The number of students from SMK Batik 1 Surakarta was 37 while from SMK Batik 2 Surakarta was 39. After the students did the critical-thinking test, the average test result obtained from the students of SMK Batik 1 Surakarta was 68.63 and that of SMK Batik 2 Surakarta was 69.17. Therefore, the developed test-items were feasible to measure students' critical-thinking skills, especially in accounting program.

## DISCUSSION

The results of this research indicated that the number of feasible test-items to measure critical-thinking skills was 6. The items were made using 4-D development model (Define, Design, Develop, Disseminate). The development model was created by Sivasaliam Thiagarajan, Dorothy Silberman Semmel, and Melvyn I. Semmel in 1974. This development model includes (1) defining stage i.e. conducting preliminary analysis, student analysis, task analysis, conceptual analysis, and determining learning objectives; (2) design stage consisting of choosing the type of prototype to be developed, designing the prototype, and creating the initial design; (3) development stage with several procedures, namely the initial design assessment of the prototype by the experts meaning that after the prototype design meets the criteria, it is tested to the students to determine the feasibility of the prototype design; (4) Dissemination stage i.e. using the prototype for wider circle.

The first step of this development model is the defining stage. According to Thiagarajan, Semmel, and Semmel (1974), this stage includes preliminary analysis, student analysis, task analysis, conceptual analysis, and determining learning objectives. The preliminary analysis in this study was to analyse the problems existing in the school including the test-items made by the teacher that had not represent critical-thinking skills. The test-items were still at the levels of remembering, understanding, and applying. These levels of thinking had not yet led to the domains of critical thinking that include analysing, evaluating, and creating. In accordance with the demands of the 21st century, critical-thinking skills are needed in the workforce (Trilling and Fadel, 2009), so that VHS students should be able to master them. In addition, VHS curriculum also says that students should be able to master critical-thinking skills (Directorate of Vocational High School Development, 2017).

Student analysis is the second analysis of the defining stage. This section describes the condition of the subjects. The subjects of the test were the eleventh-grade accounting students of VHS. They were 15-17 years old on the average. According to Piaget in Trianto (2011), students at the age of 15-17 years are able to perform activities of thinking in the abstract and can perform problem-solving activities through experiments. It is in accordance with critical-thinking skills that require problem-solving activities by abstract thinking.



Task analysis is the third part of the defining stage, containing the subject matter that must be mastered by the students. The subject matter used was adjusted to the basic competence in the financial accounting of the eleventh-grade of VHS. Directorate of Vocational High School Development (2017) states that two of the basic competencies of financial accounting are (1) Applying merchandise inventory and (2) Applying inventory-methods (FIFO, LIFO, average, and specific identification). Based on the materials formulated in the task analysis, the next step was making conceptual analysis to facilitate the understanding of subject matter. Concept analysis was made by arranging the subject matter in the form of a conceptual map as shown in chart 1 above.

Furthermore, the final part of the defining stage is the analysis of learning objectives. The making of learning objectives is in accordance with the purpose of test-items. These test-items were made to measure students' critical-thinking skills, so that the learning objectives must be in line with the purpose of the test-items. The subject matter used to make the test-items was the basic competence of applying merchandise inventory and inventory method (Directorate of Vocational High School Development, 2017). The test-items were made in accordance with the basic competencies of applying merchandise inventory and inventory methods with regard to critical-thinking indicators (analysing, evaluating, and creating). Thus, the learning objectives are (1) Students can evaluate the definition of inventory; (2) Students can distinguish inventory classification; (3) Students can distinguish merchandise inventory system; (4) Students can distinguish methods used to determine the value of inventory; (5) Students can evaluate the methods used to determine the value of inventory; (6) Students can create inventory card design.

The second step of the 4-D development model is the designing stage. Thiagarajan, Semmel, and Semmel (1974) stated that the design stage includes selecting the type of prototype to be developed, designing the prototype, from which the initial design will be formed. The prototype used in this research was the critical-thinking test-items on the basic competencies of applying inventory record and inventory method (Directorate of Vocational High School Development, 2017). The blueprint of critical-thinking test-items was made according to the outline as shown in table 3 above. The blueprint was then made into the critical-thinking test-items amounting to 14 items in accordance with the indicators of critical thinking (analysing, evaluating, and creating) (Anderson and Krathwohl, 2001).

The third step is the development stage which includes the assessment of critical-thinking test-items by experts and field trials (Thiagarajan, Semmel, and Semmel, 1974). Expert judgements and field trials were conducted to get the feasible test-items to measure students' critical-thinking skills. The test-item assessment was conducted by the expert of accounting evaluation and an accounting teacher. Based on the assessment of these experts, the theoretically feasible test-items of critical thinking were obtained. Expert judgements were based on the indicators of materials, construction, and language (Budiyono, 2015). The results showed that the material indicators scored 91.07%, the construction indicators scored 71.45%, and the language indicators scored 80.98%.

In addition, the feasibility criteria of the test-items based on the expert judgements refer to the eligibility criteria of Riduwan (2012) stating that the test-items with a score of 81% - 100% were in very-feasible category; 61% - 80% were in feasible category; 41% - 60% were in medium category; and fewer than 40% were in infeasible category. Based on the material indicators, the test-items were categorized as very-feasible items; based on the indicators of construction, the items were categorized as feasible items; and based on the language indicators, the test-items were in a very-feasible indicator. Based on the material, construction, and language aspects, the critical-thinking test-items were feasible to use in field trials.

Field trials were conducted to see the validity, reliability, discriminating power, and the difficulty levels of the test-items (Zulfadli, 2017). The result of validity test-items scoring more than 0.3044 had a valid category. According to Budiyono (2015), the determination of the validity criteria is done by comparing the value of  $r$  with the value of  $r_{table}$ . If  $r$  is greater than  $r_{table}$ , then the data is valid. If  $r$  is lower than  $r_{table}$ , then the data is not valid.  $r_{table}$  values can be searched using  $r_{table}$  in which the number of students who did the trials is 42. The next is to look for degrees of freedom (df) by using the formula of the number of test participants minus 2 ( $42 - 2 = 40$ ). Next is finding line 40 column 0.05 (5%). The  $r_{table}$  was 0.3044. The test-items with  $r$  greater than  $r_{table}$  of 0.3044 were 3, 4, 6, 7, 9, 10, 12, 13.

From the results of reliability calculations using Microsoft Excel 2013, the score obtained was 0.6548. Then, the score was compared with Alpha Cronbach of 0.6. According to Trihendradi (2012) a test-item is considered reliable if it has a score above 0.6. Therefore, the reliability value of the test-items was 0.6548 greater than 0.6, so the items of critical-thinking test were reliable. Then, the calculation results of the discriminating power were the same as the calculation of the validity level. Jihad and Haris (2008) stated that test-items having a discriminating power of  $\geq 0.40$  are in the good category; those of 0.30 - 0.39 are in the fair category; and those of 0.20 - 0.00 are in the poor category. Based on the calculation results, it was found that the test-items with good and fair discriminating power were items 3, 4, 6, 7, 9, 10, 12, and 13. Furthermore, based on the calculation of difficulty level, according to Jihad and Haris (2008), test-items with a score of 0.00 - 0.30 are in the difficult category; those with the score of 0.31 to 0.70 are in the medium category; and the items with the scores of 0.71 to 1.00 are in the easy category. The feasible test-items have a score of 0.31 - 0.70 (medium). The calculation results of the difficulty levels indicated that the medium test-items were items 1, 2, 3, 4, 7, 9, 11, 12, 13.

Based on the test results in the development stage above, the valid and reliable test-items with good and fair discriminating power and medium difficulty level were items 3, 4, 7, 9, 12, and 13. So, the number of feasible test-items to measure students' critical-thinking skills was 6. The items can be disseminated to measure students' critical-thinking skills. The last step of the 4-D development model is the dissemination stage. Thiagarajan, Semmel, and Semmel (1974) stated that the dissemination stage is using the prototype for a wider circle. Thus, the test of critical thinking was disseminated to SMK Batik 1 and SMK Batik 2 Surakarta, Central Java, Indonesia. The results of the dissemination showed that the average level of the students' critical thinking in SMK Batik 1 Surakarta was 68.63, while that in SMK Batik 2 Surakarta was 69.17.

Based on the results of the research described above, it can be concluded that the number of critical-thinking-skill test items was 14. After that, the test-items were tested to see the criteria for validity, reliability, discriminating power, and difficulty level. The results of the trial showed that there were 6 items that met the test criteria. This research is only limited to assessing test items that will be used to measure students' critical-thinking skills in accounting department of VHS. The making of the test items in this research was adjusted to the accounting subject in the eleventh grade of VHS. The suggestion for further research is to expand the making of critical-thinking test items for other subjects.

## REFERENCES

- Afsaneh, G. & Tahereh, H. (2015). A cross-contextual analysis of Iranian EFL teachers' attitudes and perceptions of critical thinking. *International Journal of Research Studies in Education*, 5 (4), 27-38.
- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Aretz, A. J., Bolen, M. T. & Devereux, K. E. (2014). Critical thinking assessment of collage students. *Journal of Collage Reading and Learning*, 28 (1), 12-23.
- Ayanwale, M. A., Adeleke, J. O. & Mamadelo, T. I. (2018). An assessment of item statistics estimates of basic education certificate examination through classical test theory and item response theory approach. *International Journal of Education Research Review*, 3 (4), 55-67.
- Budiyono. (2015). *Pengantar penilaian hasil belajar*. Surakarta: UNS Press.
- Cottrell, S. (2005). *Critical thinking skills developing effective analysis and argument*. New York: Palgrave Macmillan.
- DeWealsche, S. A. (2015). Critical thinking, questioning and student engagement in Korean University English course. *Linguistics and Education*, 32, 131-147.
- Dharmawati, Rahayu, S. & Mahanal, S. (2017). Pengembangan instrumen asesmen berpikir kritis untuk peserta didik SMP kelas VII pada materi interaksi makhluk hidup dengan lingkungannya. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 1 (8), 1598-1606.

- Saputra, M. D., Joyoatmojo, S. & Wardani, D. K. (2018). The assessment of critical thinking-skill tests for accounting students of vocational high schools. *International Journal of Educational Research Review*, 3(4), 85-96.
- Direktorat Pembinaan Sekolah Menengah Kejuruan (DPSMK) 2009 Nomor 330/D.D5/KEP/KR/2017, Kompetensi Keahlian Akuntansi dan Keuangan Lembaga.
- Dwyer, C. P., Hogan, M. J. & Stewart, I. (2014). An integrated critical thinking framework for the 21<sup>st</sup> century. *Thinking Skills and Creativity*, 12, 43-52.
- Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice*, 32 (3), 179-186.
- Epstein, R. L. (2006). *Critical thinking third edition*. Belmont: Thomson Wadsworth.
- Fatimah, A. W. N., Suryani, N. & Yamtinah, S. (2018). The development of critical thinking test based on higher-order thinking pisa version in the history learning at senior high school. *International Journal of Multicultural and Multireligious Understanding*, 5 (2), 136-144.
- Fisher, A. (2008). *Critical thinking an introduction*. Cambridge: Cambridge University Press.
- Fitzpatrick, B. & Schulz, H. (2015). Do curriculum outcomes and assessment activities in science encourage higher order thinking? *Canadian Journal of Science, Mathematics and Technology Education*, 15 (2), 136-154.
- Florea, N. N. & Hurjui, E. (2015). Critical thinking in elementary school children. *Procedia Social and Behavioral Sciences*, 180, 565 – 572.
- Hashemi, S. A., Naderi, E., Shariatmadari, A., Naraghi, M. S. & Mehrabi, M. (2010). Science production in iranian educational system by the use of critical thinking. *International Journal of Instruction*, 3 (1), 61-76.
- He, H., Craig, R. & Wen, J. (2013). Developing critical thinking skills and effective co-operative international accounting degree programs in China. *Asian Review of Accounting*, 21 (2), 144-159.
- Jihad, A. & Haris, A. (2008). *Evaluasi pembelajaran*. Yogyakarta: Multi Pressindo.
- Jones, A. (2010). Generic attributes in accounting: The significance of the disciplinary context. *Accounting Education: An International Journal*, 19, 5-21.
- Kay, K. (2008). *Preparing every child for the 21<sup>st</sup> century*. Paper presented in the APEC EdNet-Xi'an (Symposium) Xi'an China.
- Lehman, C. R. (2013). Knowing the unknowable and contested terrains in accounting. *Critical Perspectives on Accounting*, 24 (2), 136-144.
- Liu, O. L., Mao, L., Frankel, L. & Xu, J. (2016). Assessing critical thinking in higher education: The HEIghten™ Approach and Preliminary Validity Evidence. *Assessment and Evaluation in Higher Education*, 41 (5), 677-694.
- Mason, M. (2008). *Critical thinking and learning*. U.K: Blackwell.
- Mulyanto, H., Gunarhadi, & Indriayu, M. (2018). The effect of problem based learning model on student mathematics learning outcomes viewed from critical thinking skills. *International Journal of Education Research Review*, 3 (2), 37-45.
- Nahadi, Firman, H. & Kurniadi, H. (2018). Development and validation of chemistry virtual test based multiple representations. *Journal of Educational and Learning*, 12 (1), 44-51.
- Pappas, E., Pierrakos, O. & Nagel, R. (2013). Using Bloom's taxonomy to teach sustainability in multiple contexts. *Journal of Cleaner Production*, 48, 54-64.
- Paul, R. W. & Elder, L. (2002). *Critical thinking tools for taking charge of your professional and personal life*. Upper Saddle River, N. J: Pearson Prentice Hall.
- Pieterce, T., Lawrence, H. & Friedrich-Nel, H. (2016). Critical thinking ability of 3<sup>rd</sup> year radiography students. *Health SA Gesondheid*, 21, 381-390.

- Saputra, M.D., Joyoatmojo, S. & Wardani, D.K. (2018). The assessment of critical-thinking-skill tests for accounting students of vocational high schools. *International Journal of Educational Research Review*, 3(4), 85-96.
- Pradana, S. D. S., Parno. & Handayanto, S. K. (2016). Pengembangan tes kemampuan berpikir kritis pada materi optik geometri untuk mahasiswa didik fisika. *Jurnal Penelitian dan Evaluasi Pendidikan*, 21 (1), 51-64.
- Purnamawati & Saliruddin. (2017). The Effectiveness of the use of metacognitive-based industrial electronic learning tools in growing higher order thinking skills (Hots). *Jurnal Pendidikan Vokasi*, 7 (2), 139-148.
- Reinstein, A. & Bayou, M. E. (1997). Critical thinking in accounting education: Processes, skills and applications. *Managerial Auditing Journal*, 12 (7), 336-342.
- Riduwan. (2012). *Skala pengukuran variabel – variabel penelitian*. Bandung: Alfabeta.
- Saeed, K., Reza, Z. & Momene, G. (2013). Critical thinking across the elt curriculum: A mixed methods approach to analyzing L2 teachers' attitudes towards critical thinking instruction. *International Journal of Research Studies in Education*, 2 (3), 15-24.
- Syutharidho & Rosida, R. M. (2015). Pengembangan Soal Berpikir Kritis untuk Peserta didik SMP Kelas VIII. *Al-Jabar: Jurnal Pendidikan Matematika*, 6 (2), 219-227.
- Thiagarajan, S., Semmel, D. S. & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children*. Bloomington: Indiana University.
- Trianto. (2011). *Constructivist-oriented learning models*. Jakarta: Prestasi Pustaka.
- Trihendradi, C. (2012). *Step by step SPSS 20 analisis data statistik*. Yogyakarta: CV. Andi.
- Trilling, B. & Fadel, C. (2009). *21<sup>st</sup> Century skills learning for life and our times*. San Francisco: Jossey-Bass.
- Vong, S. A. & Kaewurai, W. (2017). Instructional model development to enhance critical thinking and critical thinking teaching ability of trainee students at regional teaching training center in Takeo Province, Cambodia. *Kasetsart Journal of Social Science*, 38, 88-89.
- Zubaidah, S. (2010). *Berpikir Kritis: Kemampuan Berpikir Tingkat Tinggi yang Dapat Dikembangkan melalui Pembelajaran Sains*. The Paper was Presented in the National Science Seminar 2010 with the them, "Optimizing Science to Empower Human Beings" in the Graduate School of Surabaya State University, Surabaya, East Java.
- Zulfadli. (2017). Pengembangan Tes Berbasis Revisi Taksonomi Bloom Pada Materi Struktur dan Fungsi Sel untuk Siswa Kelas XI di SMK Kota Tarakan. *Jurnal Pendidikan Biologi Indonesia*, 2 (2), 174-182.