

ASSESSING STUDENTS' HIGHER ORDER THINKING SKILLS USING MULTIDIMENSIONAL ITEM RESPONSE THEORY

Friyatmi

Padang State University & Yogyakarta State University, Indonesia
E-mail: fri.yatmi@fe.unp.ac.id

Djemari Mardapi, Haryanto

Yogyakarta State University, Indonesia
E-mail: djemariamardapi@gmail.com, haryanto@uny.ac.id

Abstract

The development of the economics HOTS test that combines critical thinking skills, problem-solving, and critical thinking are essential to meet the challenges of the 21st century life skills. The combination of these thinking skills in the HOT test will help teachers to diagnose students' strengths and weaknesses. However, it could interfere with the accuracy of the measurement results if analyzed using IRT in a single analysis. The Multidimensional Item Response Theory (MIRT) resolved the accuracy of the measurement issues. This research aimed to develop the economics HOTS test to estimate the student's abilities in higher order thinking skills using the MIRT. The samples were 750 high school students selected from fourteen high schools in West Sumatera, Indonesia. The data were collected using tests which calibrated through the simple-structure MIRT model using R studio. The test reliability was calculated based on coefficients Alpha and test information function. The results show MIRT offers accurate measurement in estimating multidimensional test parameters. The item had a moderate average multidimensional discriminant and difficulty, while the students had a moderate HOTS ability. Their ability to think creatively was lower than critical thinking and problem-solving abilities. The test was proven to be reliable with coefficients Alpha of 0.81, it yielded a high-test information function of the 4.0124, and a low measurement error of 0.4992. It is suitable to be tested on students who have moderate abilities in problem-solving and critical thinking, but with high creative thinking ability.

Keywords: HOTS, critical thinking, creative thinking, problem solving, MIRT.

Introduction

Four essential life skills that must be mastered by students in the 21st-century include the ability to think critically and problem-solving, creative and innovative, communicative, and collaborate. These skills are an indispensable underlying potential for learners to succeed in global challenges, particularly in the 4.0 revolution industrial era. The quality of human resources that can elaborate on these four life skills and master information technology is undoubtedly needed, and hence it is now a major priority for the classroom changes.

Numerous contextual teaching methods that encourage students to think critically and creatively in a variety of life contexts are optimized to help students to master the four skills. For instance, the assessment activities are directed at the mastery of these skills to test students' higher-order thinking skills (HOTS). Moreover, assessment of students' ability and cognitive knowledge are also routinely tested on an international scale like in TIMSS and PISA

(Murtiyasa, Rejeki, & Setyaningsih, 2018). Apart from those various forms of activities, the quality of teachers in the classroom remains the cornerstone for designing classroom learning and assessment, which can optimize the skill mastery.

Classroom assessment has essential role, which is the same as the learning models to encourage the acquisition of skills of learners. One of the factors which determine the quality of classroom assessment is teacher competence in designing and constructing the test items (Rahmah, 2012). If the test items only test factual knowledge, the ability of students tested is still basic and has not been able to explore the students' thinking skills. The utilization of the ability to think in learning is therefore pivotal because it can encourage intellectual growth and promote better academic achievement (Ramos, Dolipas, & Villamor, 2013). Therefore, the teacher holds a crucial role in developing the assessment that can improve thinking ability.

The teacher's ability to construct item test is prerequisite for improving the quality of learning because it might promote students' thinking ability, especially higher-order thinking skills (Sunggingwati & Nguyen, 2013). Unfortunately, considerable studies have demonstrated problems in assessment, teachers mostly test lower-order thinking skill questions, while questions that measure reasoning ability and high-level thinking are rarely administrated (Amrina, Zulkardi, & Yusuf, 2013). In general, inquiries were constructed to measure only students' ability to memorize (Fischer, Bol, & Pribesh, 2011). Undoubtedly, most teachers use lower-order thinking skills (LOTS) items for examination (Iskandar & Senam, 2015; Shidiq, Masykuri, & Van Hayus, 2014). The results have indicated a low teacher's ability in compiling HOTS items.

The teacher's lack of ability to construct HOTS questions has been shown as follows. Approximately 55% of what considered to be HOTS items written by the teacher, in fact are categorized yet as LOTS items. This research suggested that teachers were having difficulty interpreting the ability to think and make a test item for higher-level thinking.

The limited ability of teachers to construct HOTS items contributed to the students' familiarity with the factual questions which require only their memorizing ability. It is unable to encourage the ability of students to provide reasons which are in high-level thinking skills. Saido, Siraj, Nordin, and Al_Amedy (2015) tested 20 items of The Higher Order Thinking Level Test (HOTLT) to the junior high school students in Thailand, and 79.7% showed the ability of students still at the lower order thinking. This problem also happens in Indonesia, which shows that more than 50% of students are not able to solve the analysis problems, synthesize information, and make conclusions (Susanti, 2012). The results of these studies indicate that a higher level of students' ability to think in Indonesia has not been too satisfactory. A similar research also revealed only 45% of Indonesian students could solve the reasoning questions (Amirulloh, Rustaman, & Sriyati, 2014; Herman, 2007). The reasoning ability is critical in stimulating high-level thinking skills so that students can think critically about the various life problems. The results of the research have implications for the importance of the teacher's task to maximize the high-level thinking skills of students.

There have been considerable studies to develop a variety of tests to improve student HOTS, yet they still use the top-level Bloom taxonomy (Barnett & Francis, 2012; Madhuri, Kantamreddi, & Prakash Goteti, 2012; Saido et al., 2015), including in Indonesia because the model is considered superior to other taxonomies. For instance, Istiyono, Mardapi, and Suparno (2014) developed the physics HOTS test in high school using Bloom's taxonomy. The research developed multiple-choice items with polytomous scoring. Another similar research developed an essay test based on Bloom's taxonomy to measure students' higher-order thinking abilities (Lewy, Zulkardi, & Aisyah, 2009).

Bloom's taxonomy has been widely applied in the test development, but it has limitations (Amer, 2006; Booker, 2007; Hess, Jones, Carlock, & Walkup, 2009; Marzano & Kendall, 2008). Bloom's Taxonomy is more focused on learning and cognitive processes rather than

the assessment (Airasian and Miranda (2002). Therefore, other approaches in developing the HOTS test became an essential thing to be done so that the educators also comprehend other alternatives in preparing the HOTS test.

Alternative assessment to measure students' HOTS is to integrate problem-solving tests, critical thinking, and creative thinking (Haladyna, 1997; Lewis & Smith, 1993; Marzano, 1993; Patrícia, 2011). These abilities are the 21st-century life skills that should be owned by the students and relevant to the learning objectives of economics in high school, which is to identify problems and find solutions with creative and innovative thinking in real-life situations (Van Wyk, 2011). Therefore, critical thinking, problem-solving and creative thinking are needed in order to research economics so that students can produce a variety of ideas or alternative solutions needed to address the economic phenomena that occur around them.

The integration of problem-solving skills, critical thinking, and creative thinking skills in measuring HOTS has been developed by Rofiah, Aminah, and Ekawati (2013) for junior high school students in physics. However, it is using dichotomous scoring. This technique has several disadvantages because it gives a score only for the correct answer, so it cannot diagnose student errors (Isgiyanto, 2011). The latest research was conducted by Chae and Lee (2018) in measuring the HOTS ability, but it was measured using creative, critical, and caring thinking skills. Problem-solving ability is not the main focus in measuring HOTS ability in the research.

Furthermore, Budiman and Jailani (2014) developed multiple-choice items to measure math HOTS through critical and creative thinking tests in junior high school. Both of these studies have similar characteristics, which is using the same HOT indicators and using classical test theory in test analysis. The classical test theory in the test analysis offers simplicity for the users, but on the other hand, it has limitations because it is group dependent (Hambleton & Swaminathan, 2013). The nature of the group dependent classical test theory evokes dependency of item characteristics on the characteristics of participants. If an item is tested to high-level students, then the item's difficulty will be easy, otherwise, if the item is tested against participants who had low ability, then the level of difficulty of an item tends to be difficult. The weakness of classical test theory can be anticipated by applying Item Response Theory (IRT) in the test analysis.

IRT has been used by Istiyono et al. (2014) in developing physics HOTS tests for high school students. However, the underlying construct of the item test still used Bloom's taxonomy. IRT is used when a test developed is one-dimensional, where the only measure is one aspect of students' abilities (Hambleton & Swaminathan, 2013). If the IRT is used to analyze a test that measures more than one aspect of the ability, then the measurement error will occur where the item parameter estimation becomes inaccurate. The IRT limitations are fixed by Multidimensional IRT models (MIRT).

MIRT is used in the test analysis when the test is multidimensional or a test structure uses some structures that correlate with the latent variables (Reckase, 2009). Some studies show that the use of MIRT is accurate to test analysis because of its ability to estimate several different abilities in a single analysis. Ha (2016) proved MIRT more accurate than IRT in estimating English test parameters. It used three-dimensional MIRT to measure three reading competencies, namely vocab, grammar, and functions of speech.

MIRT is using principal component analysis to examine the dimensions of the test and confirmatory factor analysis to interpretation item categorization. Not only on the English test, but MIRT models are also precise in analyzing the math test (Desjardins & Bulut, 2017). The research discovered more than two-dimensional ability in math tests, specifically literacy, mathematical ability, and problem-solving. Both of these studies show that the MIRT offers more accurate measurement results and more in-depth information that test-takers can make more informed decisions regarding test results.

Thus, the development of the economic HOTS test that combines critical thinking skills, problem-solving, and critical thinking are essential to implement because these are needed to meet the challenges of the 21st century life skills. The mixture of these thinking skills is in the HOT test which will help teachers to diagnose students' strengths and weaknesses at every level. The three abilities are interrelated with one another, and limited numbers of research investigating this realm, the analysis test using MIRT therefore matters. MIRT is an accurate analysis tool for measuring the inter-related ability in the HOTS tests.

Research Problem

The utilization of thinking skills is very important because it can encourage intellectual growth in learning. One way to improve thinking skills is through assessment in the form of questions that test higher-order thinking skills. Higher-order thinking skills are important abilities that students must master in facing the challenges of life in the 21st century, especially in economics teaching.

There are many approaches that are commonly used to measure higher-order thinking skills however, Bloom's taxonomy has been a priority in various research. It offers convenience but is more focused on learning than assessment. Some important skills such as critical thinking, problem solving, and creative thinking are less explored when using Bloom's taxonomy. In this research, HOTS ability was measured using critical thinking skills, problem-solving, and creative thinking. The selection of the three approaches is relevant to the abilities required in economics learning and in accordance with the minimum thinking abilities students must master in facing the challenges of the 21st century life skills. The combination of the three skills in one test will cause problems if the parameters are estimated in a single analysis using IRT. MIRT offers an accurate estimate for test analysis that consists of several correlated dimensions. The HOTS Economics test was developed based on good test development procedures and was calibrated using MIRT to produce valid items, reliable instruments, and to produce a more accurate estimation of students' ability. Therefore, the formulation of the problem raised in this research is how is the parameter estimated of the Economic HOTS test when analyzed using MIRT?

Research Focus

The main focus of this research was to assess students' higher order thinking skills through the HOTS Economics test. The objectives of this research are as follows: 1) to validate higher-order thinking skills test on economics (henceforth Eco-HOTS test), 2) to assess the item parameter of Eco-HOTS test empirically, 3) to assess the student's HOTS ability, 4) to reveal the Eco-HOTS test parameter.

Research Methodology

General Background

Research and development method was employed in this research. It is used to develop the Eco-HOTS test to measure students' higher order thinking skills. Test development consisted of developing test specifications, item development, assembly test, field testing, analysis, and revision. It was adapted from the procedure of test development (Downing, 2011; Oriondo & Dallo-Ontonio, 1998). The research was conducted in the senior high schools in West Sumatra province, Indonesia from July to November 2018.

Sample

The sample was 750 students of the 11th grade, comprising 308 males and 442 females. It was taken by proportional random sampling technique from fourteen public and private senior high schools located in the cities and suburbs of West Sumatera Province, Indonesia. The students were selected based on their variation of abilities, taken from schools with low, medium, and high categories on national exam scores. The difference in school areas is assumed to be contributed to the inequality of school quality and student achievement. The data were obtained from the West Sumatera National Examination Scores of 2017 published by The Ministry of National Education.

Instrument and Procedures

The instrument in this research was Eco-HOTS test developed by the researcher. It consisted of three constructs to measure economic higher order thinking skills, namely critical thinking, problem solving, and creative thinking skills. The Eco-HOTS test was constructed using a test specification by adapting Bransford's IDEAL problem-solving test, the California Critical Thinking Skills Test (CCTT), and the Torrance Test Creative Thinking (TTCT). Based on the specification test then the Eco-HOTS test items were advanced using multiple-choice, constructed response, and essay format. The test package contained twenty items consisting of seven items of critical thinking, ten items of problem-solving, and three items of creative thinking. Creative thinking skills are tested using an essay test to explore the various solutions offered by students, while critical thinking and problem-solving skills are tested with multiple-choice questions and constructed-response tests. All items used polytomous scoring with a score of 1-4.

Test items were validated using content validity through expert judgment. Five experts assessed the accuracy of each item in measuring higher-order thinking skills by giving a score of 1 to 4. It was collected using item validation sheet. Items that have good content validity were chosen to be assembled into test packages. Furthermore, a test was conducted to determine students' abilities in higher-order thinking skills. It is also carried out to estimate the psychometrical properties of the items and test parameters.

Data Analysis

The content validity of the items test was validated by the Aiken formula. Aiken formulated the Aiken's V formula to calculate the content-validity index based on the results of the judgment of an expert panel of n people on an item in terms of the extent to which the item represented the measured construct.

$$V = \frac{\sum s}{n(c - 1)}$$

V is Aiken's item validity index, s is the score given by rater minus the lowest score in the rating, n is the number of raters, and c number of rating (Azwar, 2012).

Items' parameter and students' ability were estimated using simple-structure MIRT. The MIRT procedure begins with testing the test dimensions and then continues with parameter estimation. The dimensionality testing was analyzed using the factor analysis method (Finch, 2006) with regard to the scree plot and eigenvalues. If there are several components that have more than one eigenvalue on the screen plot, then it proves that the test has a multidimensional trait.

Common models used in simple-structure MIRT include multidimensional graded response model (M-GRM), multidimensional partial credit model (M-PCM), and bi-factor models. The item parameter of the M-GRM generates parameters a and d . The a parameter was a multidimensional vector of the discriminant power that can be interpreted as the slope in the IRT models. The d parameter was a scalar associated with the difficulty of items or also called intercept or location in the IRT. Both of these parameters are fundamental in determining the level of multidimensional difficulty (MDIFF) and multidimensional discriminant (MDISC).

$$MDISC = \sqrt{\sum_{k=1}^m a_{ik}^2}$$

a_i is a vector of a discriminant parameter, and m is number of dimensions (Ackerman, Gierl, & Walker, 2003)

$$MDIFF_{ik} = \frac{-d_{ik}}{MDISC}$$

d is scalar of difficulty parameter, and MDISC is items' multidimensional discriminant index (Reckase, 2009).

The interpretation of the MIRT parameters is similar to the IRT. Discriminant index of a good item is 0 to 2, and the item difficulty is ranging from -3 to +3 (Hambleton, Swaminathan, & Rogers, 1991). The ability parameter (θ) lies in the interval $-\infty$ to $+\infty$ and is scaled close to the normal distribution with mean 0 and standard deviation 1. In practice, ability lies between -3 to +3 (Brennan, 2006). Students' abilities are grouped into three ability levels, i.e., high, moderate, and low abilities. It is categorized using the mean ideal score.

Table 1
Ability category

Interval score	Category
$X > Mi + 1,5 SBi$	High
$Mi - 1,5 SBi < X \leq Mi + 1,5 SBi$	Moderate
$X \leq Mi - 1,5 SBi$	Low

Reliability tests based on classical test theory approach can be seen from the value of reliability, according to the item response theory known from information function tests. Test information function is the sum of the whole grain items information functions at the level of ability (Hambleton, Swaminathan, & Rogers, 1991). The value of the item information functions is estimated using equation below.

$$I_{\alpha}(\theta) = \frac{[\nabla_{\alpha} P_i(\theta)]^2}{P_i(\theta) Q_i(\theta)}$$

α is a vector of coordinate axes x , which gives the direction of centroid θ . ∇_{α} is directions derivative α . $P_i(\theta)$ is the probability of answering correctly, and $Q_i(\theta) = 1 - P_i(\theta)$. Based on this formula, the total test information was calculated by summing all of the item information.

The values of the test parameters are the estimation results so that the truth is a probability and cannot be separated from measurement errors.

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (4)$$

SEM (q) is a standard error measurement, and I (q) is test information.

Research Results

The Eco-HOTS test contains 20 items that measure three dimensions of thinking. Item specifications for each dimension are described in Table 2.

Table 2
Test specifications

Dimension	Sub-skills	Number of items	Format item
Problem-solving	Identify the problem Define the problem Examine the options Act on a plan Look at the consequences	7	Multiple choices and constructed response
Critical thinking	Interpretation Analysis Inference Evaluation	10	Multiple choices and constructed response
Creative thinking	Fluency Flexibility Originality	3	Essay

Item Validation

The validity of the items was analyzed using the content validation index Aiken. The decision on the validity of the items carried out by comparing the Aikens' index for each item with a reference value of Table Aiken. The reference value based on Aikens' Table for five raters and four ratings was 0.87. Each item was valid if it had 0.87 or more of Aikens' index. The calculations of the twenty items were validated and showed that 20 items were categorized as valid items because they had an Aiken index ≥ 0.87 . The results of the Aikens' index for each item are shown in Table 3.

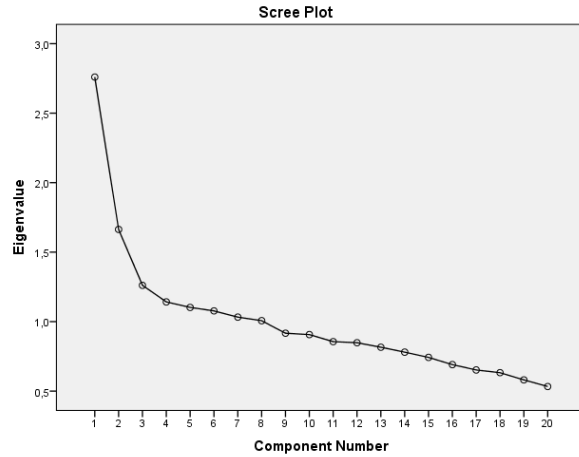
Table 3
Items content validity

Item	Σs	n (c -1)	V	Results
1	13	15	0.87	valid
2	13	15	0.87	valid
3	14	15	0.93	valid
4	13	15	0.87	valid
5	14	15	0.93	valid
6	14	15	0.93	valid
7	14	15	0.93	valid
8	14	15	0.93	valid
9	14	15	0.93	valid
10	14	15	0.93	valid
11	15	15	1.00	valid
12	13	15	0.87	valid
13	14	15	0.93	valid
14	14	15	0.93	valid
15	14	15	0.93	valid
16	14	15	0.93	valid
17	13	15	0.87	valid
18	15	15	1.00	valid
19	13	15	0.87	valid
20	14	15	0.93	valid

Dimensionality Testing

The dimensionality test was performed to analyze the latent variables involved in determining the score of each item. Dimensionality testing in this research was not aimed to explore the number of dimensions because it was determined using test specifications. Dimensionality testing was proved using factor analysis. It aimed to determine the multidimensional trait of the Eco-HOTS test, not to determine the number of dimensions involved in the test. The dimensionality testing results delineated in Figure 1.

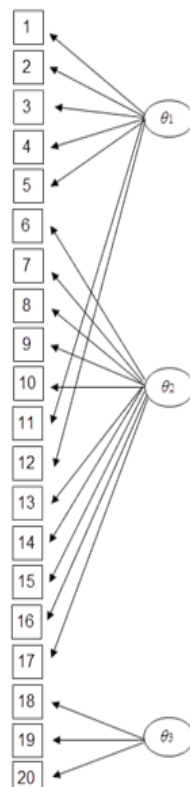
Figure 1
Scree plot for dimensionality testing



The scree plot in Figure 1 indicates that several factors have eigenvalue over 1. It means several factors play a role in determining the students' abilities. In other words, it can be said that the Eco-HOTS test had a multidimensional trait. Based on these results, the researchers decided to categorize the Eco-HOTS test items in their respective latent trait according to the initial design.

The latent ability attributes of the Eco-HOTS test (θ) consisted of critical thinking skills (θ_1), problem-solving ability (θ_2), and creative thinking skills (θ_3). The structure of the Eco-HOTS test used a simple-structure MIRT, which means that one item only measures one ability. The correlation of items with latent traits is illustrated through the test dimension structure according to Figure 2.

Figure 2
Structure dimensions of the Eco-HOTS test



Items' Parameters

Item parameters were analyzed using M-GRM. It is used for polytomous score items and each category is ordered. Estimation of parameters using M-GRM had two parameters, namely the parameters *a* and *d*. The number of parameters *a* depends on the number of dimensions involved in a test, while the number of parameters *d* for polytomous scoring can be presented as an overall difficulty or categorical difficulty. The results of the item parameters of the Eco-HOTS tests are presented in Table 4.

Table 4
MIRT items parameter

Item	a1	a2	a3	d1	d2	d3
Item1	1.61	0	0	0.53	-0.29	-1.52
Item2	0.93	0	0	2.37	1.20	-0.12
Item3	0.87	0	0	3.62	1.37	-1.81
Item4	1.59	0	0	1.83	0.92	-1.09
Item5	0.96	0	0	3.46	0.33	-1.88
Item11	0.46	0	0	4.03	0.48	-0.81
Item12	0.59	0	0	3.40	1.14	-1.33
Item6	0	0.55	0	4.20	0.77	-2.38
Item7	0	0.34	0	4.27	1.04	-0.52
Item8	0	0.98	0	4.73	1.23	-3.40
Item9	0	0.83	0	4.54	0.83	-2.03
Item10	0	0.26	0	2.99	0.57	-0.74
Item13	0	0.58	0	1.65	0.30	-0.31
Item14	0	0.68	0	2.51	0.16	-1.03
Item15	0	-0.21	0	0.94	0.41	-0.77
Item16	0	2.40	0	4.49	-0.40	-1.22
Item17	0	2.25	0	3.63	0.02	-0.78
Item18	0	0	1.46	1.82	-3.34	-5.06
Item19	0	0	2.30	1.80	-4.75	-8.45
Item20	0	0	0.45	-0.24	-3.79	-6.30

Parameters a1, a2, a3 relate to the slope for each dimension ability, i.e. a1 is slope for critical thinking items, a2 is a slope of problem-solving items, and a3 is a slope for creative thinking items. Since the Eco-HOTS test is using simple model-structure MIRT, then the *a* parameter will only be available to the dimensions set, whereas a value for the other dimension will be worth 0. Table 2 shows that the first seven items relate to slope parameters of critical thinking, the next ten items are problem-solving, and three last items are creative thinking.

The *d* parameter is associated with location probabilities. Students were able to answer correctly 50% of an item. Parameters d1, d2, d3 is a categorical location related to the level of

difficulty of the test scores for the polytomous scoring. A positive value indicates a student's chances to answer more than 50% correctly, while a negative value was vice versa.

Table 3 shows that the d1 have the positive index which means that the chances of students to answer category/score one is over than 50%. While the d3 have dominant-negative value, which means the probabilities of students to answer precisely category three is less than 50%. It indicates that it is difficult for students to get the score 3.

The parameters a and d cannot be regarded as multidimensional difficulty and discriminant of items. Therefore, the conversion parameters a and d must be multidimensional discriminant (MDISC), and multidimensional difficulty (MDIFF) is needed in order to provide information related to the difficulty and discriminant index. The results of the MDISC and MDIFF estimation of the Eco-HOTS test are described in Table 5.

Table 5
Item multidimensional difficulty and discriminant index

Dimension	No. Item	MDISC	MDIFF
Critical Thinking	Item1	1.61	0.26
	Item2	0.93	-1.23
	Item3	0.87	-1.22
	Item4	1.59	-0.35
	Item5	0.96	-0.66
	Item11	0.46	-2.69
	Item12	0.59	-1.80
Problem Solving	Item6	0.55	-1.57
	Item7	0.34	-4.72
	Item8	0.98	-0.87
	Item9	0.83	-1.34
	Item10	0.26	-3.54
	Item13	0.58	-0.94
	Item14	0.68	-0.80
Creative Thinking	Item15	0.21	-0.91
	Item16	2.40	-0.40
	Item17	2.25	-0.43
	Item18	1.46	1.50
	Item19	2.30	1.65
	Item20	0.45	7.64

The results show that there is only one MDISC value in Table 5 that represents the items discriminant for the dimensions measured according to the simple-structure MIRT. Seventeen items have a good discriminant power because the value is in the range of 0-2, while three items (items 16, 17, 19) have a high discriminant index. The MDIFF value is conferred as the overall item difficulty level, which is the average of the categorical difficulty. The multidimensional difficulty index of 3 items (items 11, 7, 10) are easy, 1 item (item 20) is difficult, and the remaining 16 items have moderate multidimensional difficulty index. Three easy items measure critical thinking skills and problem-solving, while one difficult item measures creative thinking skills.

The mean of multidimensional discriminant for each HOTS dimension had a moderate index and the mean of multidimensional difficulty for critical thinking and problem-solving dimension had a moderate difficulty index, while the creative thinking had a high multidimensional difficulty index. It indicated that creative thinking test was more difficult than critical thinking and problem-solving. Overall, the multidimensional difficulty index of the Eco-HOTS test had a mean of 0.32, and a mean of the multidimensional discriminant index was 1.11. Both item parameters were in the moderate category.

Based on the results, the Eco-HOTS test had a good multidimensional difficulty, characterized by its dominant items in moderate difficulty and only a little bit of items in the difficult and easy category. A comparison between the three HOTS dimensions exposed to the multidimensional difficulty of creative thinking skills had a high difficulty index than the other HOTS skills.

Students' Parameter

Students' abilities were in the range -2.23 to 2.20. It is a good ability because it is in the range of -3 to 3. The mean ability for each HOTS dimension is shown in Table 6.

Table 6
Ability estimation

Dimension	Mean (\bar{X})	SD	Max	Min
Critical Thinking	0.0009	0.69	1.76	-1.59
Problem Solving	-0.0025	0.58	2.03	-1.41
Creative Thinking	-0.0059	0.78	2.20	-2.23
Grand Mean	-0.0028			

The result representing the overall mean of students' ability was -0.0028. It is a moderate category because it was close to 0. The three-dimensional MIRT results show the students' ability to have a moderate ability for each HOTS dimension. The highest student's ability was a critical thinking ability, whereas creative thinking ability was lower than critical thinking and problem-solving. It indicates that students' skills to provide various solutions to economic problems are still limited in economics learning. The category of students' abilities is classified in Table 7.

Table 7
Percentages of student ability category

HOTS Dimension	High	Moderate	Low
Critical Thinking	8	81	11
Problem Solving	14	83	3
Creative Thinking	2	80	18

Most students have had moderate abilities for each HOTS skill. The number of students who have high ability in creative and critical thinking was less than the number of students who have low ability. Conversely, the number of students who have high ability in problem-solving was more than the low category.

Test Parameters

Measurement accuracy is shown by the test reliability and test information function. The test has had a high reliability of 0.81. Thus, it can be regarded as a reliable test. It has had a maximum test information function of the 4.0124 with a measurement error of 0.4992. The results indicated that the measurement results provide a high information function with low measurement errors.

The test provides maximum information on the participants' ability $\theta = 0$ for critical thinking and problem solving as well as $\theta = 2$ for creative thinking. It means that the test was suitable to be tested on students who have moderate abilities in problem-solving and critical thinking, but with high creative thinking ability.

Discussion

There were three HOTS dimensions measured by the test, critical thinking, problem-solving, and creative thinking skills. Comparison of students' ability equalization on the three aspects of HOTS shows that students' abilities in critical thinking were higher than problem-solving and creative thinking. These results indicate that students' critical thinking skills were better than others.

The finding is consistent with Hendricson et al. (2006) which revealed that problem-solving is a late manifestation of critical thinking skills. Therefore, it is common that the problem-solving ability is lower than critical thinking because they have no solution to offer. Since students already have high skills of critical thinking, then the teacher can facilitate the improvement of students' problem-solving abilities by providing materials discussion or task-based problem-solving. This finding is consistent with the studies of Tmkaya, Aybek, and Aldađ (2009) and Kkdemir (2003) that revealed students who have high levels of critical thinking skills that will encourage problem-solving skills and decision-making. If students are accustomed to practice in solving the problems, their creative thinking abilities will also be honed.

The results prove that among the three dimensions, the creative thinking skills have had a higher difficulty index than problem-solving and critical thinking. Supposedly, there are several factors that cause it.

First, creative thinking items require the student to provide various solutions to the problems presented. The ability to produce a solution is not an easy activity, but it requires complex thinking skills and even needs the ability to imagine. Unfortunately, not everyone has a high imaginative ability as far more accustomed to thinking systematically. Teaching methods and environments that do not support the development of creative thinking ability are thought to contribute to affect it. Some research show that the selection of appropriate learning strategies and educational environment impacts on student creativity (Gregory, Hardiman, Yarmolinskaya, Rinne, & Limb, 2013; Horng & Lee, 2009). Moreover, creating a learning environment that encourages students to learn is a challenge that is not easy for teachers in the 21st century (Tsciak-Deliowska, 2018). The research results indicate the importance of the role of teachers in facilitating students' ability to think creatively.

Second, the most challenging aspects of creative thinking ability are to provide original ideas, new and unique ideas. Questions of the test which contain original ideas are a tough

question because the ability of high school students is generally not being able to provide the original idea. Most of the students are only able to offer ordinary ideas even tend to be textbook based, i.e. solutions that have already been provided on economics books.

Third, Students' HOTS ability demonstrates that the abilities of problem-solving and creative thinking were in the middle category. Research of Istiyono, Dwandaru, and Faizah (2018) showed that it is not too satisfying students' skills in problem-solving, of which about 36% of students have moderate ability and other students have the low capacity in problem-solving. Based on the findings of this research, the selection of instructional design and assessment that can optimize students' problem-solving ability is essentially implemented for students accustomed to thinking critically and able to provide various solutions to problems encountered. Although students' creative thinking and problem-solving were in the middle category, creative thinking abilities tend to be lower than others. One strategy to encourage the students' problem-solving abilities is by presenting questions that raise an issue in the cultural context or local customs (Hamdi, Suganda, & Hayati, 2018). This approach was also used by the researchers to develop one of the Eco-HOTS items, particularly by presenting customs of Indonesian society in the face of Ramadhan fasting and Eid and its relationship to inflation.

Simple application in economics assessment, for example for the inflation concept, when students are asked to mention any policy can be issued by the government to tackle the problem of inflation, the students are not too difficult to mention other forms of monetary and fiscal policy to cope with inflation. However, when students were asked to provide simple ways they can do to deal with the problem of inflation in everyday life, many students found difficult to answer this question. This is reflected in the responses of the test they replied. Many students are not able to provide answers to these questions.

Even so, there were also some students who gave interesting ideas. For instance, some students offer alternative investments in the form of game tokens and forex trading. The first idea may seem illogical; nonetheless, the idea is given quite unique because the student has earned money from playing games. The second idea is a high-level idea for a high school student because not everyone has an understanding of trading. Moreover, the topic has never been discussed by the teacher.

Once confirmed to the students, the first student is a student who likes to play games and often earns money from selling game tokens. The second student has parents who participate in forex trading, so he gets some information about trading from his parents. These findings prove that students can offer solutions in their own creative ways of thinking. It is one of the effective approaches stimulating students to provide solutions through creative thinking and imagination (Gonda & Tirpakova, 2018).

Testing the high order thinking skills can also be applied to the math concept in economics, for instance, on the inflation topic in high school. If the matter were made merely to ask students to calculate inflation, then the case is that the students can easily answer it. Teachers can restructure the basis of the questions so that more students' thinking skills increase. The input is presented in the following two forms of matter.

Question A.

Consumer Price Index (CPI) for a commodity in 2016 of IDR 10000.00 and 2017 IDR 9800.00. Calculate the rate of inflation in 2017!

Question B.

If the Consumer Price Index (CPI) for a commodity of IDR 10.000.00 in 2016 and 2017 amounted to IDR 9800.00, does it really reflect inflation of 2%?

Question A is a form of matter that is very common in school and national test for inflation topic. This item still measures the application's ability that is still classified as lower order thinking skills. Question B is similar to question A, but the student should have to analyze the problem to answer the item correctly. Question B has a higher cognitive level than question A, which involves the HOTS item. Question B is one of the items tested on the Eco-HOTS test. Unfortunately, many students are not able to answer the question correctly, because so far students are accustomed to being given application questions, especially calculating the inflation rate. It indicates that students are still having trouble answering HOTS questions because they are not familiar. Students' unfamiliarity with HOTS questions can be one of the causes of students' difficulties in solving HOTS items (Hadi, Retnawati, Munadi, Apino, & Wulandari, 2018).

Based on interviews conducted with some students, they stated that the HOTS questions were something new for them. Teachers rarely even test HOTS questions on exams. It is justified by the teacher who states that HOTS problems are rarely tested on students all this time. Even some teachers said they found difficulty to answer the Eco-HOTS test. The Eco-HOTS test can be a reference for teachers to develop HOTS questions on different topics in economics learning or other subjects. These questions are then validated and calibrated, so they can be submitted in the regional item banks that have been developed previously (Friyatmi, Mardapi, Haryanto, & Rahmi, 2020). When the item bank contains HOTS questions, then students' abilities can be tested on classroom assessment continuously and systematically. Therefore, it is very important to promote HOTS questions to teacher and high school students. The high school age is a critical period for the future, and the creative and innovative ability is expressed in this age (Mikhailova, 2018). Assessing students with HOTS items will be able to improve students' thinking skills, especially the ability to think creatively.

Conclusions and Implications

Based on the results, it can be concluded that MIRT offers accurate measurement results in estimating multidimensional test parameters. It establishes the estimation of three dimensions of HOTS abilities at once in a single analysis with small measurement errors. Such precise results cannot be obtained if IRT is used as a tool to estimate test parameters, because it is only able to estimate one dimension in one analysis.

A comparison between HOTS dimensions showed that the ability of the students in creative thinking was lower than critical thinking and problem-solving. Moreover, the Eco-HOTS test was reliable and suitable to be tested on students who have a moderate ability in critical thinking and problem solving, but it has a high ability in creative thinking.

The results have implications for the amelioration of the teaching learning by improving the students' ability of creative thinking in learning. It could be implemented by using authentic assessment by presenting real-life issues related to the social context and everyday life problems.

Acknowledgments

The authors would like to offer their gratitude to the Ministry of Research, Technology and Higher Education of Indonesia for the research grant.

References

- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, 22(3), 37-51. <https://doi.org/10.1111/j.1745-3992.2003.tb00136.x>
- Airasian, P. W., & Miranda, H. (2002). The role of assessment in the revised taxonomy. *Theory into Practice*, 41(4), 249-254. https://doi.org/10.1207/s15430421tip4104_8
- Amer, A. (2006). Reflections on Bloom's revised taxonomy. *Electronic Journal of Research in Educational Psychology*, 4(8), 213-230.
- Amirulloh, D., Rustaman, N., & Sriyati, S. (2014). Analisis soal SNMPTN biologi berdasarkan domain kognitif taksonomi bloom revisi dan profil capaian siswa SMA kelas XII [Analysis of SNMPTN biology item based on the Bloom revised cognitive taxonomy domain and the 12th grade high school students' achievement profile]. *Formica Education Online*, 1(1), 1-8.
- Amrina, R., Zulkardi, & Yusuf, H. (2013). Pengembangan soal penalaran model TIMSS matematika SMP [Development of mathematical reasoning questions like TIMSS for junior high school]. *Jurnal Penelitian dan Evaluasi Pendidikan*, 17(2), 230 -240.
- Azwar, S. (2012). *Reliabilitas dan validitas* [Reliability and validity]. Pustaka Pelajar.
- Barnett, J. E., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, 32(2), 201-211. <https://doi.org/10.1080/01443410.2011.638619>
- Booker, M. J. (2007). A roof without walls: Benjamin Bloom's taxonomy and the misdirection of American education. *Academic Questions*, 20(4), 347-355. <https://doi.org/10.1007/s12129-007-9031-9>
- Brennan, R. (2006). *Educational measurement*. Greenwood Publishing Group.
- Budiman, A., & Jailani, J. (2014). Pengembangan instrumen asesmen higher order thinking skill (HOTS) pada mata pelajaran matematika SMP kelas VIII semester 1 [Development of assessment instruments for higher order thinking skills (HOTS) in mathematics for 8th grade junior high school]. *Jurnal Riset Pendidikan Matematika*, 1(2), 139-151.
- Chae, S. E., & Lee, M.-S. (2018). Determinants of latent profiles in higher-order thinking skills of Korean university students. *Problems of Education in the 21st Century*, 76(4), 483-498.
- Desjardins, C. D., & Bulut, O. (2017). *Handbook of Educational Measurement and Psychometrics Using R*. CRC Press.
- Downing, S. M. (2011). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 1-28). Lawrence Erlbaum Associates.
- Finch, H. (2006). Comparison of the performance of varimax and promax rotations: Factor structure recovery for dichotomous items. *Journal of Educational Measurement*, 43(1), 39-52. <https://doi.org/10.1111/j.1745-3984.2006.00003.x>
- Fischer, C., Bol, L., & Pribesh, S. (2011). An investigation of higher-order thinking skills in smaller learning community social studies classrooms. *American Secondary Education*, 39(2), 5-26.
- Friyatmi, Mardapi, D., Haryanto, & Rahmi, E. (2020). The development of computerized economics item banking for classroom and school-based assessment *European Journal of Educational Research*, 9(1), 293-303. <https://doi.org/10.12973/eu-jer.9.1.293>
- Gonda, D., & Tirpakova, A. (2018). A new teaching method aimed at eliminating the causes of students' unsuccessful algorithmic problem solving with parameter. *Problems of Education in the 21st Century*, 76(4), 499-519.
- Gregory, E., Hardiman, M., Yarmolinskaya, J., Rinne, L., & Limb, C. (2013). Building creative thinking in the classroom: From research to practice. *International Journal of Educational Research*, 62, 43-50. <https://doi.org/10.1016/j.ijer.2013.06.003>
- Ha, D. T. (2016). Applying multidimensional item response theory in validating an English final test. *Journal of Technical Education Science*, 6(36), 103-110
- Hadi, S., Retnawati, H., Munadi, S., Apino, E., & Wulandari, N. F. (2018). The difficulties of high school students in solving higher-order thinking skills problems. *Problems of Education in the 21st Century*, 76(4), 520-532.
- Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Allyn and Bacon.
- Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publication.
- Hamdi, S., Suganda, I. A., & Hayati, N. (2018). Developing higher-order thinking skill (HOTS) test instrument using Lombok local cultures as contexts for junior secondary school mathematics. *Research and Evaluation in Education, 4*(2), 126-135.
- Hendricson, W. D., Andrieu, S. C., Chadwick, D. G., Chmar, J. E., Cole, J. R., George, M. C., Haden, N. K. (2006). Educational strategies associated with development of problem-solving, critical thinking, and self-directed learning. *Journal of Dental Education, 70*(9), 925-936.
- Herman, T. (2007). Pembelajaran berbasis masalah untuk meningkatkan kemampuan penalaran matematis siswa SMP [Problem-based learning to improve mathematical reasoning abilities of junior high school students]. *Jurnal Cakrawala Pendidikan, 1*(1), 41-62.
- Hess, K. K., Jones, B. S., Carlock, D., & Walkup, J. R. (2009). Cognitive rigor: Blending the strengths of Bloom's taxonomy and Webb's depth of knowledge to enhance classroom-level processes. *ERIC Online Submission, 1*(1), 1-8.
- Horng, J.-S., & Lee, Y.-C. (2009). What environmental factors influence creative culinary studies? *International Journal of Contemporary Hospitality Management, 21*(1), 100-117. <https://doi.org/10.1108/09596110910930214>
- Isgiyanto, A. (2011). Analisis data ujian nasional matematika berdasarkan penskoran model rasch dan partial credit [Analysis of national mathematics examination data based on the Rasch model and partial credit]. In *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA*. Yogyakarta, Indonesia.
- Iskandar, D., & Senam, S. (2015). Studi kemampuan guru kimia sma lulusan UNY dalam mengembangkan soal UAS berbasis HOTS [Study of the chemistry teacher ability who graduated of UNY in developing HOTS-based UAS]. *Jurnal Inovasi Pendidikan IPA, 1*(1), 65-72.
- Istiyono, E., Dwandaru, W. S. B., & Faizah, R. (2018). Mapping of physics problem-solving skills of senior high school students using PhysProSS-CAT. *Research and Evaluation in Education, 4*(2), 144-154.
- Istiyono, E., Mardapi, D., & Suparno, S. (2014). Pengembangan tes kemampuan berpikir tingkat tinggi fisika (pysthots) peserta didik SMA [Development of high-level physics thinking skills tests (pysthots) for high school students]. *Jurnal Penelitian dan Evaluasi Pendidikan, 18*(1), 1-12.
- Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Journal Theory into Practice, 32*(3), 131-137. <https://doi.org/10.1080/00405849309543588>
- Lewy, L., Zulkardi, Z., & Aisyah, N. (2009). Pengembangan soal untuk mengukur kemampuan berpikir tingkat tinggi pokok bahasan barisan dan deret bilangan di kelas IX akselerasi SMP xaverius maria palembang [Test development to measure the high order thinking skills on topic line and number sequence in class IX acceleration xaverius maria junior high school palembang]. *Jurnal Pendidikan Matematika, 3*(2), 14-28.
- Madhuri, G., Kantamreddi, V., & Prakash Goteti, L. (2012). Promoting higher order thinking skills using inquiry-based learning. *European Journal of Engineering Education, 37*(2), 117-123. <https://doi.org/10.1080/03043797.2012.661701>
- Marzano, R. J. (1993). How classroom teachers approach the teaching of thinking. *Theory into Practice, 32*(3), 154-160. <https://doi.org/10.1080/00405849309543591>
- Marzano, R. J., & Kendall, J. S. (2008). *Designing and assessing educational objectives: Applying the new taxonomy*. Corwin Press.
- Mikhailova, O. B. (2018). Features of creativity and innovation development in students at secondary and high school. *International Journal of Cognitive Research in Science, Engineering and Education, 6*(2), 11-20. <https://doi.org/10.5937/ijersee1802011M>
- Murtiyasa, B., Rejeki, S., & Setyaningsih, R. (2018). PISA-like problems using Indonesian contexts. In *Journal of Physics: Conference Series, 1040*(1), 1-8. <https://doi.org/10.1088/1742-6596/1040/1/012032>
- Oriondo, L. L., & Dallo-Ontonio, E. M. (1998). *Evaluating educational outcomes*. REX Book Store.
- Patricia, A.-A. (2011). Critical thinking, questioning and creativity as components of intelligence. *Procedia-Social and Behavioral Sciences, 30*(1), 357-362. <https://doi.org/10.1016/j.sbspro.2011.10.070>
- Rahmah, J. (2012). Domain soal PISA untuk literasi matematika. *Jurnal Peluang, 1*(1), 30-42.

- Ramos, J. L. S., Dolipas, B. B., & Villamor, B. B. (2013). Higher order thinking skills and academic performance in physics of college students: A regression analysis. *International Journal of Innovative Interdisciplinary Research*, 1(4), 48-60.
- Reckase, M. D. (2009). *Multidimensional item response theory models*. Springer.
- Rofiah, E., Aminah, N. S., & Ekawati, E. Y. (2013). Penyusunan instrumen tes kemampuan berpikir tingkat tinggi fisika pada siswa SMP [Developing of high-order thinking physics test instruments for junior high school students]. *Jurnal Pendidikan Fisika*, 1(2), 17-22.
- Saido, G. M., Siraj, S., Nordin, A. B. B., & Al_Amedy, O. S. (2015). Higher Order Thinking Skills Among Secondary School Students in Science Learning. *The Malaysian Online Journal of Educational Science*, 3(3), 13-20.
- Shidiq, A. S., Masykuri, M., & Van Hayus, E. S. (2014). Pengembangan instrumen penilaian two-tier multiple choice untuk mengukur keterampilan berpikir tingkat tinggi (higher order thinking skills) pada materi kelarutan dan hasil kali kelarutan untuk siswa SMA/MA kelas XI [Development of two-tier multiple choice assessment instruments to measure higher order thinking skills on solubility material and solubility results for grade XI high school/MA students]. *Jurnal Pendidikan Kimia*, 3(4), 83-92.
- Sunggingwati, D., & Nguyen, H. T. M. (2013). Teachers' questioning in reading lessons: A case study in Indonesia. *Electronic Journal of Foreign Language Teaching*, 10(1), 80-95.
- Susanti, E. (2012). Profil higher order thinking skills dan mathematical habits of mind siswa: Studi kasus pada siswa sekolah menengah atas untuk topik statistika [Profile of higher order thinking skills and students' mathematical habits of mind: Case studies in high school students for statistical topics]. *Forum MIPA*, 15(2), 120-127.
- Tłuściak-Deliowska, A. (2018). About the school and the student-teacher relationship in the 21st century: Some perspectives and challenges. *Problems of Education in the 21st Century*, 76(4), 422-424.
- Tümkiye, S., Aybek, B., & Aldağ, H. (2009). An investigation of university students' critical thinking disposition and perceived problem-solving skills. *Eurasian Journal of Educational Research*, 36, 57-74.
- Van Wyk, M. M. (2011). The effects of Teams-Games-Tournaments on achievement, retention, and attitudes of economics education students. *Journal Social Science*, 26(3), 183-193. <https://doi.org/10.1080/09718923.2011.11892895>

Received: *December 03, 2019*

Accepted: *March 06, 2020*

Cite as: Friyatmi, Mardapi, D., & Haryanto (2020). Assessing students' higher order thinking skills using multidimensional item response theory. *Problems of Education in the 21st Century*, 78(2), 196-214. <https://doi.org/10.33225/pec/20.78.196>

Friyatmi (Corresponding author)	Assistant Professor, Lecturer and Researcher at Padang State University and PhD Student at Yogyakarta Sate University, Jl. Kolombo Karangmalang No. 1 Yogyakarta 55281, Indonesia. E-mail: fri.yatmi@fe.unp.ac.id Website: http://fe.unp.ac.id/faculty/friyatmi https://scholar.google.com/citations?hl=en&user=yr_7uXIAAAAJ ORCID: https://orcid.org/0000-0001-7361-233X
Djemari Mardapi	Professor, Lecturer and Researcher, at Yogyakarta Sate University, Jl. Kolombo Karangmalang No. 1 Yogyakarta 55281, Indonesia. E-mail: djemarimardapi@gmail.com Website: http://staff.uny.ac.id/dosen/prof-djemari-mardapi-mpd-phd https://scholar.google.com/citations?user=nEtsTFwAAAAJ&hl=en&oi=ao ORCID: https://orcid.org/0000-0002-2656-3544
Haryanto	Doctor, Lecturer and Researcher, Student at Yogyakarta Sate University, Jl. Kolombo Karangmalang No. 1 Yogyakarta 55281, Indonesia. E-mail: haryanto@uny.ac.id Website: http://staff.uny.ac.id/dosen/dr-haryanto-mpd-mt