Development of Higher-Order Thinking Skills Assessment for High School Economic Subject

Dadang Dahlan[1✉]

[1]Faculty of Business and Economic Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

Article Info	Abstract
	Along with adopting a new curriculum in schools, there has been a shift of approach in terms of assessment matter that requires an assessment to be more comprehensive. This study aims to identify the economics teachers' competency deficit in developing the instruments for measuring higher-order thinking skills. The research method is Research and Development; it is a method to generate certain products and test product effectiveness. The product is in the form of an instrument model. The participants are 45 economics teachers and 270 students. The result shows that teachers' competency in making questions that involve a higher cognitive level is still considered low. It is found that only 12% of the teachers' questions are at the range of cognitive level analyzing to level creating, while the rest 88% are at the range of cognitive level remembering to level applying. This finding reflects the teachers' competency deficit in drafting HOTS questions based on the higher-order thinking aspect. According to this finding, an instrument model was then developed to measure higher-order thinking skills (HOTS). The developed instrument models are in the form of both multiple-choice and essay tests.

✉ Correspondence :
Faculty of Business and Economic Education, Universitas
Pendidikan Indonesia, Bandung, Indonesia
E-mail: dadangdahlan@ymail.com

INTRODUCTION

Instructional management is an activity to plan to teach and learning programs, implement the teaching and learning process, and assess the process and learning outcomes and the development of classroom management (Danarwati, 2016). In this matter, the main tasks of professional educators are to educate, teach, guide, direct, assess, and evaluate the students in formal childhood education school, primary, and secondary schools. As an evaluator, a teacher must collect, analyzes, interpret and finally make a judgment over the success rate of the learning process based on the established criteria with regards to the effectiveness of the process and product qualifications.

According to the 2013 Curriculum, the skills that must be assessed by teachers must include higher-order thinking skills. The students must master thinking skills from a low level (Lower Order Thinking Skills - LOTS) to a higher level (Higher Order Thinking Skills - HOTS). To have higher-order thinking skills is essential because an intelligent and skilled generation is characterized by the ability to think at a higher level (higher-order thinking skills) in solving complex problems faced in life. Generations with higher-order thinking skills will be able to compete in the era of the highly competitive nature of globalization nowadays.

The data from the result of Trends in International Mathematics and Science (TIMSS) survey in 2011 show that Indonesian students' thinking skill is considered low (Mullis, Martin, Foy, & Arora, 2012). Indonesian students can only work on questions that require low thinking skills (questions that only require skills such as knowing or recall).

Based on a World Bank report, which the study was done by the International Association for the Evaluation of Educational Achievement (IEA)in Asia showed that the reading skill of Indonesian Fourth graders of primary schools was at the lowest position of the ranking. The average reading test score of Indonesian students is at 51.7 only, meaning that Indonesian students can only master 30% of the reading materials and are still having difficulty in answering questions in essay form, which requires higher-order thinking skills. Students in Indonesia are accustomed to answering questions made by teachers that are at a low cognitive level.

The low ability of Indonesian students in terms of the application of higher thinking shows that the quality of education in Indonesia still needs to be improved in many ways, including the improvement of teachers' competencies. One of the teachers' competencies that should be improved is in drafting questions to measure higher-order thinking skills.

The low quality of questions made by the teachers is shown by several indicators, such as invalid and unreliable, no test blueprint, weak in linguistic aspect, and measure low cognitive only. As described by the researchers that: Research has shown that teachers lack sufficient training in test development, fail to analyze tests, do not establish reliability or validity, do not use a test blueprint, weight all content equally, rarely test above the basic knowledge level, and use tests with grammatical and spelling errors. (Haynie, 1992).

In line with the results of research done by the Balitbang Ministry of National Education, which express the low quality of the teachers-made tests, "most of the questions have not yet met all the steps or rules that must be followed when drafting multiple-choice questions, so in other ords, the questions made by the teacher are not at the maximum quality" (Ekaningrum, 2002).

In reference to the writer of Module of Drafting higher order thinking skills questions (Kemdikbud, 2017), HOTS questions are instruments to measure higher-order thinking skills, thinking skills that are not just recalled, restate, or referring without further process/recite. HOTS questions in the context of assessment measure skills include: 1) transfers one concept from one to another, 2) process and apply information, 3) search for connection from different information received, 4) use information to solve problems and 5) scrutinizes ideas and information critically. Even so, HOTS questions are not always more complicated than recalling questions. HOTS questions also measure the meta-cognitive dimension, and not only

measuring factual, conceptual, or procedural dimensions. The meta-cognitive dimension describes skills to connect several different concepts, interpreting, problem-solving, selecting a strategy to solve problems, discover new methods, reasoning, make the right decisions.

Furthermore, it can be described that drafting HOTS questions generally by using stimuli. Stimuli are the base in making questions. In the context of HOTS, the presented stimuli should be contextual and exciting. Stimuli may be sourced from global issues such as information technology, science, economy, health, education, and infrastructure problems. Other than stimuli utilization, HOTS questions generally measure skills such as analyzing (*analyzing*-C4), evaluating (*evaluating*-C5), and creating (*creating*-C6). In selecting the operating verbs (KKO) to formulate the HOTS questions indicators, it is suggested to do not be trapped in KKO grouping. For example, the verb 'determine' in Taxonomy Bloom belongs to C2 (comprehension) and C3 (application). In the context of writing HOTS questions, the verb 'determine' shall be categorized as C5 (evaluating) in the case of the decision-making process is preceded by a thinking process of analyzing the information at stimuli and finally, the students are asked to make the best decision. Even the verb 'decide' shall be categorized as C6 (creating) if the questions demand constructing strategy skills as new problem-solving. In conclusion, operating verbs are greatly affected by the necessary thinking process to answer the given questions.

Along with curriculum updates in many countries, empirical studies in association with learning and assessment based on higher-order thinking skills (HOTS) were already conducted in many countries, including Malaysia. In Malaysia, learning related to higher-order thinking skills has already begun in the early 1990s. It is reflected in one of the objectives of higher education in Malaysia, which reads to develop and enhance their (students') intellectual capacity concerning rational, critical, and creative thinking.

One study was conducted in Australia by Nesbitt (2005). This research was supported by the fact that the world shifts so rapidly that the students are expected to develop higher-order thinking skills in accomplishing their school works in every subject. And so are with the Australian students that are expected to use various higher-order thinking skills at every subject in school. In a science class in Queensland, higher-order thinking skills are assessed from the complexity of logical thinking and the scientific process's skills. The study reviewed how well the students in the science class utilized higher-order thinking skills when using computer simulations in accomplishing their complex science tasks. Quantitative and qualitative methods were deployed in this research through a case study. The results of the study show that simulations encourage students to think more deeply about the subject material and debate in confidence and in the right way. Students' competence in argumentation increased; students were also satisfied with the obtained learning outcomes. Students study science with their awareness and discover the knowledge needed by them.

A study in Indonesia that is in harmony with the implementation of 2013 curriculum renewal was exampled by Istiyono, Mardapi, and Suparno (2015), who conducted the research to develop instruments for Physics' higher-order thinking skills (PhysTHOTS) for High School students and to obtain the characteristics of PhysTHOTS. The instrument's blueprint was constructed based on aspects and sub-aspects of higher-order thinking skills, which then utilized to develop items of questions. The instrument consisted of two test kits, and each contained 26 items, including eight anchor items and validated by the experts in measurement, in physics education, and physics practitioners. The valid instrument was then tried on the students in 10 High School in the Special Region of Yogyakarta. Polytomous data were analyzed using the Partial Credit Model (PCM) analysis. The trial result showed that 44 items were compatible with the PCM, so the PhysTHOTS instrument was proved to be compatible with PCM; with instrument reliability of 0.95, difficulty index was between -0.86 and 1.06, which means all items in the good category. Thus, PhysTHOTS meets the requirements to measure the Physics higher-order thinking skills of high school students.

One of the emphasized aspects is that assessment regarding cognitive areas should stress more to assess Higher Order Thinking Skills (HOTS). This study is aimed at (1) to identify the competency deficit of the economics teachers in developing the instruments for measuring higher-order thinking skills, (2) to generate an instrument model for measuring higher-order thinking skills of economics students in High School. Creating an assessment as an instrument model, it was expected that it might facilitate the teachers to develop questions for measuring higher-order thinking skills. Other than that, it may be used to train the teachers for learning assessment. The produced assessment instrument model is also useful to train the prospective teachers from the university to have sufficient competency in evaluating learning outcomes.

RESEARCH METHODS

The selected research method is research and development; it is a method to generate a certain product and to test the product effectiveness in the form of instrument models. This kind of research belongs to research and development, according to Creswell (2017). Likewise, Gall, Gall, & Borg. (2003) set limits on research and development to develop and validate products used in education. The Research subjects consist of economics teachers and students at High Schools (SMAN) who have experienced the implementation of the 2013 curriculum. The participants on the first step of the research consisted of 45 economics teachers in 27 High School in Bandung City and students with a total of 270 people. In the second step, the research subject consisted of economics teachers and students with a total of 45 and 64, respectively, of whom were determined purposively and proportionally.

Tests were deployed as the main instrument tried out to the students to test higher-order thinking skills. Documentation was made to analyze questions made by the teachers so that the competence in constructing questions can be measured furthermore, while the questionnaire was deployed to know the students' responses toward questions that were tried out.

After data was collected, the data was then analyzed through quantitative analysis using descriptive statistics. To process data to test validity, reliability, difficulty levels, and discrimination index, software Anates software was deployed.

RESULTS AND DISCUSSION

The analysis result of the questions made by the teachers shows that the teachers' skills in making questions at a cognitive level are still low. The calculation result shows that only 12% of the teachers-made questions that are at the cognitive level analyzing to level creating, while the other 88% is at the cognitive level remembering to level applying. This finding reflects a teachers' competency deficit in drafting questions based on higher-order thinking skills aspect. The teachers' competency deficit in drafting questions is not only in the matter of cognitive skills but also in drafting instruments (making questions) based on contextual problems. Based on the analyzed document, only 10% of the questions are categorized as contextual-based problems.

Following the design, an instrument model has been developed consisting of: (1) contextual problems-based essay questions and (2) multiple-choice questions. The blueprint is developed within Basic Competencies in class X: (1) Describing the formation of market equilibrium and market structure, (2) Presenting observation on changes in price and quantity of equilibrium in the market. The selection of these Basic Competencies is carried out with consideration that the Basic Competencies are essential in economics.

The developed model was then analyzed quantitatively and qualitatively). Qualitative analysis is done to study both the construction and linguistic aspects. Quantitative analysis was done after the instrument was tried-out to the students from the two schools, in 11 High School Bandung = 33 students and 15 High School Bandung = 31 students, hence the total number of students participated as try-out subjects are = 64 students.

The try-outs were done in class X after finishing the learning in Basic Competencies, as explained above. The try-out results from the two schools are presented in Table 1.

Table1. Result of Validity Test on Students of 11 High School Bandungand15 High School Bandung

Question	Types of Question	r_{xy}	r_{Table}	Information
1	Multiple Choice	0.175	0.250	Invalid
2	Multiple Choice	0.085	0.250	Invalid
3	Multiple Choice	0.445	0.250	Valid
4	Multiple Choice	0.144	0.250	Invalid
5	Multiple Choice	0.485	0.250	Valid
6	Multiple Choice	0.218	0.250	Invalid
7	Multiple Choice	0.345	0.250	Valid
8	Multiple Choice	0.569	0.250	Valid
9	Multiple Choice	0.414	0.250	Valid
10	Multiple Choice	0.408	0.250	Valid
11	Multiple Choice	0.071	0.250	Invalid
12	Multiple Choice	0.537	0.250	Valid
13	Multiple Choice	0.152	0.250	Invalid
14	Multiple Choice	0.635	0.250	Valid
15	Multiple Choice	0.505	0.250	Valid
16	Essay	0.903	0.250	Valid
17	Essay	0.903	0.250	Valid

The data presented in Table 1 showed that out of 17 questions that were tested, there were 6 invalid questions they are questions number 1, 2, 4, 6, 11, and 13, all of them are multiple-choice questions. On the other hand, essay questions had a higher validity degree than multiple-choice questions. The try-out results concerning reliability are presented in Table 2.

Table2. The Result of Reliability Test on Students of 11 High School and 15 High School in Bandung

Question	Types of Question	Reliability of estimate	Information
1 – 15	Multiple Choice	0.94	Reliable
16 – 17	Essay	0.77	Reliable

The data that is presented in Table 2 showed that according to the try-out result, both multiple-choice and essay questions have a good reliability index score.

Other than validity and reliability, by doing the try-outs, the difficulty index and Discrimination index of the questions were also acknowledged. Data with regards to the Difficulty Index are presented in Table 3.

Table 3. Test Result of Difficulty Index on the Questions Tested to Students of 11 High School and 15 High School Bandung

Question	Types of Question	Difficulty Index	Information
1	Multiple Choice	0.172	Difficult
2	Multiple Choice	0.750	Easy
3	Multiple Choice	0.453	Medium
4	Multiple Choice	0.875	Easy
5	Multiple Choice	0.484	Medium

Question	Types of Question	Difficulty Index	Information
6	Multiple Choice	0.891	Easy
7	Multiple Choice	0.141	Difficult
8	Multiple Choice	0.297	Difficult
9	Multiple Choice	0.625	Medium
10	Multiple Choice	0.344	Medium
11	Multiple Choice	0.188	Difficult
12	Multiple Choice	0.641	Medium
13	Multiple Choice	0.234	Difficult
14	Multiple Choice	0.547	Medium
15	Multiple Choice	0.531	Medium
16	Essay	0.300	Medium
17	Essay	0.310	Medium
	Mean	0.478	

From the data presented in Table 3, there are 5 items of question that are categorized as Easy. They are questions number 2, 4, and 6. While questions number 1, 7, 8, 11, and 13 are considered Difficult. Furthermore, the questions that are categorized as Medium there are as many as 9 in total. They are numbers 3, 5, 9, 10, 12, 14, 15, 16, and 17. For essay questions, the Difficulty Index is in the Medium category. According to the analysis result of the classical test theory approach, the tried-out package of questions has an average of the Difficulty Index of 0.478. From the average difficulty indexes, the questions were in a Good category.

Table 4. Test Result of Discrimination Index on Questions Tested to Students of 11 High School Bandung and 15 High School Bandung

No	Types of Question	Discrimination Index	Information
1	Multiple Choice	0.17	Bad
2	Multiple Choice	0.10	Bad
3	Multiple Choice	0.44	Good
4	Multiple Choice	0.14	Bad
5	Multiple Choice	0.49	Good
6	Multiple Choice	0.22	Bad
7	Multiple Choice	0.35	Fair
8	Multiple Choice	0.57	Good
9	Multiple Choice	0.41	Good
10	Multiple Choice	0.40	Good
11	Multiple Choice	0.07	Bad
12	Multiple Choice	0.53	Good
13	Multiple Choice	0.15	Bad
14	Multiple Choice	0.63	Good
15	Multiple Choice	0.50	Good
16	Essay	0.38	Fair
17	Essay	0.40	Fair
	Mean	0.30	

Classically, the Discrimination index is -1.00 to +1.00. According to data presented in Table 5.4, there are as many as nine items of question that have a Discrimination index of more

than 0.3. They are questions number 3, 5, 7, 8, 9, 10, 12, 14, 15, 16, and 17. While the other six items of the question have a Discrimination index of less than 0.3, they are questions number 1, 2, 4, 6, 11, and 13. The average of the Discrimination index is 0.302, which means the averages of questions are in a Good category.

According to data obtained from the students through a questionnaire that was distributed previously after the try-out, it is described that the responses gained from the students from two schools are relatively different. As many as 70% of the students who participated in the try-out in High School 11 Bandung responded that the tested questions were at Difficulty Index, and as many as 80% of the students commented that to answer the questions in the try-out, some Higher Order Thinking Skills (*HOTS*) were required. This happened as the students were not accustomed to doing the tried-out questions. In this matter, as many as 76% of the students admitted that they were not accustomed to doing such questions. On the contrary, the students in High School 15 Bandung responded differently as 79% of the students there confessed that the given questions' difficulty is at the Medium category, and only 21% of the students who participated in the try-out who said that the questions were at Difficult category. Then, as many as 55% of the students responded that answering the questions require higher-order thinking skills (*HOTS*). There are only 62% of the students who responded that they were not accustomed to answering the tried out questions. According to the data above, the students' responses in High School 15 Bandung are more moderate and considered more congruent with the instrument try-out result. The try-out result showed that only 6 questions (35%) that are considered difficult. The result of data analysis showed that the average test score of the students in High School 15 Bandung is far better than it is in High School 11 Bandung. The students in both schools are in harmony said that in answering such questions some Higher Order Thinking Skills were required (*HOTS*).

Discussion

The try-out results were checked to see if the items meet the criteria, reliability, difficulty index, discrimination index, and validity score. A good reliability coefficient is indicated by a score of more than 0.70 (Linn, 1989; Mardapi, 2014). The data presented in Table 2 showed that according to the try-out results, seen from the coefficient of reliability of the questions, both the multiple-choice and essay type questions have a good reliability index coefficient. A good Difficulty Index in classic theory is when a question has a medium Difficulty Index, it is 0.3 to 0.7 (Nitko, 1983). For essay, the Difficulty Index is at a medium category. Based on the results of the analysis of classical theories, the tried-out package of questions had an average Difficulty Index of 0.478. When viewed from the average Difficulty Index, the questions were indicated as category Good.

Discrimination index is classically from -1,00 to +1,00. Not only Reynolds, Livingston & Wilson (2010) but also Kartowagiran (2012) suggest that the accepted items of questions are those with the Discrimination index of more than 0.3. According to the data presented in Table 5.4, as many as nine items of the question have a Discrimination index score of more than 0.3; they are numbers 3, 5, 7, 8, 9, 10, 12, 14, 15, 16, and 17. On the other hand, the Discrimination index score for the other six questions is less than 0.3. They are numbers 1, 2, 4, 6, 11, and 13. The average index score of Discrimination is at 0.302, which means under category Good. Furthermore, the validity is checked, and it was found that six questions are invalid needs revision. They are numbers 1, 2, 4, 6, 11, and 13. A revision was made to improve questions to meet the requirements.

CONCLUSION

Teachers' competency in drafting questions at the cognitive level is still considered low. The calculation result showed that only 12% of the teachers-made questions that are at a cognitive level range from level analyzing to level creating, while other 88% of the questions are at a cognitive

level remembering to level applying. This finding reflects a competency deficit in developing a HOTS instrument. The developed instrument models are in the form of multiple-choice and essay questions. Based on the quantitative analysis of the try-out results, the questions had passed the reliability test, Difficulty Index, and Discrimination power. For the six invalid questions, revision and improvements were made.

It is necessary to design an effective training model to enhance teachers' competency in doing an assessment, especially the HOTS-based assessment model. Therefore, collaboration among parties is needed (Department of Education) to facilitate effective training about test assessment.

REFERENCES

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Danarwati, Y. S (2016), *Manajemen Pembelajaran dalam Upaya Meningkatkan Mutu Pendidikan*, Retrieved Desember 26,2016, from http://stiaasmisolo.ac.id/jurnal.php/jmbb/article/download

Ekaningrum, T.N. (2002). *Hubungan antara Pengetahuan Standardisasi Tes Prestasi Belajar dan Sikap terhadap Profesi Guru dengan Kualitas Soal Buatan Guru Matematika di Lingkungan MGMP SLTP DKI Jakarta (2001)*. Tesis Magister pada PPS UNJ: tidak diterbitkan.

Gall, M.D., Gall, J.P., & Borg, W.R. (2003). *Educational Research. An Introduction*. Seventh Edition. Boston: Pearson Education, Inc.

Haynie, W.J. (1992). "Post Hoc Analysis of Test Items Written by Technology Education Teachers". *Journal of Technology Education. 4*(1), 26-38.

Istiyono, E., Mardapi, D., & Suparno. (2014). "Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PhysTHOTS) Peserta Didik SMA". *Jurnal Penelitian dan Evaluasi Pendidikan. 18*(1), 1-12.

Kartowagiran, B. (2012). *Penulisan Butir Soal*.MakalahdisajikandalamPelatihan Penulisandan Analisis Butir SoalbagiSumberdaya PNS Dik-Rekinpeg, di Hotel KawanuaAerotel, Jakarta.

Kemdikbud. (2017). *Modul Penyusunan Soal Higher Order Thinking*. Jakarta: Direktorat Pembinaan SMA.

Linn, R.L. (1989). *Education Measurement (3rd ed.)*. New York: MacMillan Publishing Company.

Mardapi, D. (2014). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Yogyakarta: NuhaLitera.

Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics* (pp. 139-171). Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Nesbitt, P. J. (2005). *Higher Order Thinking Skills in a Science Classroom Computer Simulation*. A Thesis at Queensland University of Technology Brisbane.

Nitko, A. (1983). *Educational Test and Measurement: An Introduction*. New York: Harcourt Brace Jovanovich, Inc.

Rajendra, N. (2001). "The Teaching of Higher-Order Thinking Skills in Malaysia "*Journal of Southeast Asian Education 2*(1), 42-65

Reynolds, C. R., Livingston, R. B., & Wilson, V. (2009).*Measurement and Assessment in Education (2nd ed)*. Boston: Pearson Education Inc.

Sangadji, E.M. & Sopiah. (2010). *Metodologi Penelitian-Pendekatan Praktis dalam Penelitian.* Yogyakarta: Penerbit Andi.