JOURNAL INTELLECTUAL SUFISM RESEARCH (JISR)
e-ISSN: 2622-2175
p-ISSN: 2621-0592
JISR 4(1), November 2021, 22-30
Email: journaljisr@gmail.com

Development of Higher-order Thinking Skills-Based Test Instruments on Internasional Trade Materials

Risma Ananda Putri[1], Dhiah Fitrayati[2].
[1]Economic Education, Universitas Negeri Surabaya, Indonesia
[2]Economic Education, Universitas Negeri Surabaya, Indonesia

Abstract. This research is developing a test instrument based on higher-order thinking skills on international trade material. The objectives of this development are 1) to develop a test instrument based on higher-order thinking skills; 2) measure the validity and reliability of the test instrument; 3) identify distinguishing power and level of difficulty of the test instrument; and 4) measuring the effectiveness of the test instrument distractors. This development uses the Borg & Gall development model consisting of potentials and problems, data collection, product design, design validation, design revision, revision collection, product revision, use trial, and product revision. The results showed that the 10 questions were valid and reliable with the result that the validity value of each question was more than 0.338 and the reliability was 0.655. From the results show that the questions includes characteristics of higher-order thinking skills. The test items already fulfilled the requirements of good test, but still need some revision on some items.

Keywords: *Higher-order thinking skills, Tests, Critical Thinking*

INTRODUCTION

The era of education in the 21st-century is experiencing rapid development due to the development of information and communication technology. 21st-century education plays an important role in forming quality students with 21st-century skills (Wijaya et al., 2016). In the concept of 21st-century skills is the term 21st-century knowledge-skill rainbow which has three core subjects, namely life and career skills (life and career skills), learning and innovation skills (learning and innovation skills), and communication media and technology skills (information media and skills). technology skills) (Trilling & Fadel, 2009). These skills prepare students to face life in the era of globalization (Iberahim et al., 2017).

One of the subjects in the 21st-century knowledge-skill rainbow is learning and innovation which has three aspects, namely (1) critical thinking and problem solving (expert thinking); (2) communication and collaboration (complex communication); and (3) creativity and innovation (applied invention and imagination) (Trilling & Fadel, 2009). National Education Association (2002) suggests that "The 4Cs" as 21st-century skills consist of critical thinking, communication, collaboration, and creativity. Although there are different opinions about the distribution of the number of skill points, they both have the same core purpose. Meanwhile, Iberahim et al (2017)

Corresponding author: Risma Ananda Putri. Email: rismaanandaputri36@gmail.com

Doi: https://doi.org/10.52032/jisr.v4i1.101

said that one of the important soft skills for students for 21st-century education is critical and structured thinking, and problem solving skills.

The concept of critical thinking and problem solving is very important concept because education is not only about learning, teaching, or learning outcomes but also how to think (Marniwati, 2019). According Syutharidho & M (2015), critical thinking is a valid, significant, natural, and accurate mindset in solving problems. While higher-order Thinking Skills are thinking skills to produce decisions based on logical and scientific reasons (Supranoto, 2018).

Higher-order thinking skills according to Kemendikbud (2017) is an assessment tool used as a measuring tool for higher-order thinking skills above remembering, restating, or referring without being processed. A question is categorized as a higher-order thinking skills question if the question is categorized as C4 (analyzing) and C5 (creating) cognitive levels (Supranoto, 2018). Meanwhile, to make good higher-order thinking skills questions according to Uno & Koni (2013), they must meet indicators, namely problem solving, creative thinking skills, and decision-making skills.

The results of the 2018 Program for International Student Assessment (PISA) survey on performance in science, reading, and mathematics. Of the 79 countries that participated in the PISA survey, Indonesia was ranked 74th. The total score obtained by Indonesia on science ability was 396, reading 371, and mathematics 379, this score is far below the OECD average score (Schleicher, 2018). Indonesia's score is at level 1 with a score range of 340-440, from this score it is known that Indonesia's education level is still at the stage of completing tasks with low complexity, limited collaboration skills and still need instructions in taking action (OECD, 2015).

Similar results are also seen in the results of the National Examination for the 2018/2019 academic year, namely that many students answered questions correctly on the low cognitive shutter, but on the contrary on the high cognitive shutter from C4 to C6 the majority of students could not solve the exam questions. From the results of the percentage of students who answered correctly, it is known that the lowest percentage is in the basic competence of international trade, especially the material on currency exchange, the percentage at the provincial level is 18.12% and the national level is 22.01% and the balance sheet material with the percentage at the provincial level is 15.62. % and the national level is 16.75%.

Based on the PISA survey, the distribution of competency achievements for the 2018/2019 National Examination and the importance of critical thinking in 21st-century learning, students' critical thinking skills need to be trained. One way to train students' critical thinking skills is through the habit of solving questions based on higher-order thinking skills (Suhady et al., 2020). This is because in analyzing higher-order thinking skills questions, a critical, reflective, logical, creative, and metacognitive reasoning process is needed (Lewy et al., 2009). Thus, when students succeed in solving higher-order thinking skills based questions, then the student is considered to have been able to improve higher-order thinking skills (Oktanisa & Fitrayati, 2018).

This study aims to 1) develop a test instrument based on higher-order thinking skills; 2) measure the validity and reliability of the test instrument; 3) identify distinguishing power and level of difficulty of the test instrument; and 4) measuring the effectiveness of the test instrument distractors. The research output is a valid and reliable higher-order thinking skills-based test instrument. The output of this research is expected to be able to help educators to formulate higher-order thinking skills questions and to be able to test students' critical thinking levels.

METHOD

This research is a research and development (R&D) development model of Borg & Gall. The Borg & Gall development model consists of 10 steps, namely potential and problems, data collection, product design, design validation, design revision, collection of revisions, product revisions, usage trials, product revisions, and mass production (Sugiyono, 2017). For the last stage, mass production was not carried out due to the limitations of researchers in terms of time and cost.

The initial stage of this development is to identify potential problems for students. The second stage is collecting data about the curriculum used and the characteristics of the students to be studied. After receiving data from the school, it was continued with the third stage, namely the design of the product of the question which started from determining the purpose of the test, compiling a grid of questions, reviewing questions, testing questions, and assembling questions. Assembled questions will later be validated by an evaluation expert in the fourth stage, then if there are deficiencies in the product, revisions are made, revisions are returned, and product revisions are still found in the fifth, sixth, and seventh stages. After the revision stage is complete, it enters the eighth stage, namely the questions are ready to be tested on students to test the validity, reliability, discriminating power, level of difficulty, and the effectiveness of the test instrument distractors. If in the trial phase there are still invalid questions, a revision will be made again at the ninth stage.

This study uses qualitative and quantitative data. The results of the test data by the validator will be used as qualitative data, while the data from the product analysis including validity, reliability, discriminating power, level of difficulty, and effectiveness of distractors will be included in quantitative data.

Test the validity of the items using product moment correlation analysis. Items are declared valid if the value of r count > r table (Kadir, 2015). While the reliability test of the question uses Cronbach's Alpha analysis. Items are declared reliable if the value of Cronbach's Alpha ≥ 0.60 and ≤ 1 (Arifin & Retnawati, 2017).

Test the discriminatory power of items using the AnatesV4 application. The distinguishing power of items is categorized into 5 groups, namely:

Table 1. Different Power Criteria

Differing Power index	Criteria
1,00 – 0,70	Very good
0,70 – 0,40	Goog
0,40 – 0,20	Enough
0,20 – 0,00	Not good
Negative	Question not used

Arikunto (2009).

To measure the level of difficulty of the items using the analysis of the average value (mean). The level of difficulty of the questions is categorized into 3 categories, namely:

Table 2. Difficulty Level Criteria

Mean	Criteria
$0,7 < p$	Easy
$0,70 \geq p \geq 0,3$	Medium
$0,3 > p$	Hard

Supranoto (2018).

The distractor effectiveness analysis was carried out using the AnatesV4 application. The effectiveness of distractors is measured by the pattern of answers chosen by students from the available alternative answers (Uno & Koni, 2013). The effectiveness of the item distractors is categorized into 5 groups, namely:

Table 3. Distractors Effectiveness Criteria

Category	Distractor Index	Criteria
- - -	> 200%	Very bad
- -	200% - 176%	Bad
-	175% - 151%	Not good
+	150% - 126%	Good
++	125% - 76%	Very good

Arifin & Retnawati (2017).

RESULT AND DISCUSSION

The development of the test instrument uses the Borg and Gall development stages. In the first stage, a potential and problem analysis is carried out. Based on the results of the 2018 PISA survey on performance in science, reading, and math skills from 79 participating countries, Indonesia was ranked 74th which indicates that the ability of Indonesian students still does not meet PISA standards. According to Suhady et al (2020) students can develop thinking skills through higher-order thinking skills problem solving exercises.

The second stage is information gathering. Based on observations at SMAN 1 Sidoarjo, it is known that the school uses the 2013 curriculum as a teaching guide. Assessment activities are

carried out using Mid-Semester Assessment (PTS), Final Semester Assessment (PAS), and Year-End Assessment (PAT). The types of questions used are objective and subjective questions with a proportion of more objective questions. Judging from the aspect of thinking ability according to Bloom's taxonomy, most of the questions are at levels C1, C2, C3, and some levels C4. While the questions discussed in the learning activities are sourced from Student Worksheets (LKS). The results of the study of the items contained in the LKS show that most of the questions presented are at the C3 level. Thus, that student learning activities are less trained in solving hots-based questions. According to (Supranoto, 2018), questions based on higher-order thinking skills can to think at least at the C4 level.

Based on the information obtained in the second stage, the development of a test instrument based on higher-order thinking skills was carried out. Based on the results of the study of the material contained in the 2013 curriculum and the existing syllabus, the material chosen for the development of a higher-order thinking skills-based test instrument is international trade material. The selection of the material is based on the results of the 2018/2019 school year national exam, in which the achievement level is low. After the material is determined, the next step is to review the study material in international trade material. The study of study materials was carried out to identify study materials and formulate indicators according to level 4 thinking skills in Bloom's taxonomy. Based on the results of the study materials, 10 indicator formulations were obtained which would later be developed into items based on higher-order thinking skills.

Judging from the type of questions, the items developed were designed in the form of multiple choice. The question body design consists of a stimulus and a subject matter. Stimulus questions present phenomena and/or contextual data in the form of graphics/images, visualizations, text, etc. While the subject matter is formulated using operational verbs that reflect the ability to think at least C4 in Bloom's taxonomy. Both the stimulus and the subject matter are developed based on the indicators that have been formulated. Technically, the items were developed by referring to the guidelines for the preparation of items based on higher-order thinking skills issued by the Ministry of Education and Culture.

If the questions have been developed, then the next step is to validate the items to the validator. Item validation includes aspects of material, construction, and language. The items that have been validated by the validator are revised and re-validated until the items are declared feasible and ready to be tested on students. In this study, item validation was carried out 5 times.

Items that have been declared feasible by the validator then enter the trial stage. The test phase of the questions was tested on 32 students of class XI IPS 3 at SMAN 1 Sidoarjo. Furthermore, the output of the test questions will be analyzed in terms of quantitative including validity, reliability, discriminating power, level of difficulty, and effectiveness of distractors. Testing the data using the IBM SPSS statistics version 23 application to test the validity, reliability, and difficulty level. Meanwhile, to test the distinguishing power and effectiveness of the distractors, the anatesV4 application was used.

Items are declared valid if the value of r count > r table (Kadir, 2015). The results of the validity test can be seen in table 4. Based on the table, it is known that all items are declared valid with 1 question having high validity, 7 questions having sufficient validity, and 2 questions having low validity.

Table 4. Validity Test Results

Number	R table	R count	Criteria
1	0,338	0,504	Enough
2	0,338	0,370	Low
3	0,338	0,517	Enough
4	0,338	0,627	High
5	0,338	0,543	Enough
6	0,338	0,358	Low
7	0,338	0,517	Enough
8	0,338	0,416	Enough
9	0,338	0,523	Enough
10	0,338	0,544	Enough

An item is declared to meet the reliability requirements if the value of Cronbach's Alpha ≥ 0.60 and ≤ 1 (Arifin & Retnawati, 2017). The results of the analysis using SPSS show the Cronbach's Alpha value of 0.655, so the question is reliable.

The test of distinguishing power of questions is carried out to determine the ability of students to work on the questions (Uno & Koni, 2013). Different test analysis was performed using the AnatesV4 application. The categories of discriminating power of items are grouped into 4 categories (see table 1). The results of the analysis of the distinguishing power index are known to all the questions have a good category as presented in table 5.

Table 5. Different Power Test Results

Number	Different Power Indeks	Criteria
1	55,56	Good
2	44,44	Good
3	66,67	Good
4	66,67	Good
5	66,67	Good
6	44,44	Good
7	66,47	Good
8	44,44	Good
9	66,67	Good

10	66,67	Good

The level of difficulty of the items can be known through the results of the average value (mean) in the statistical table (Kartowagiran, 2012). The level of difficulty of the items is categorized into 3 categories (see table 2). The results of the analysis show that 9 out of 10 items have a difficulty level in the "Medium" category with a mean value interval of $0.3 \leq p \leq 0.7$ and 1 item has a difficulty level in the "Hard" category because the mean value is < 0.30.

Table 6. Difficulty Level Test Results

Number	Mean	Criteria
1	0,69	Medium
2	0,31	Medium
3	0,56	Medium
4	0,50	Medium
5	0,56	Medium
6	0,63	Medium
7	0,56	Medium
8	0,28	Hard
9	0,50	Medium
10	0,59	Medium

The distractor effectiveness analysis was conducted to determine the quality of the distractors by using the AnatesV4 application. The category of item distractor effectiveness is grouped into 5 categories (see table 3). The results of the analysis of the effectiveness of the item distractors are presented in table 7.

Table 7. Distractor effectiveness test

Number	Total distractor	Criteria
- - -	10	Very good
- -	12	Good
-	13	Not bad
+	3	Bad
++	2	Very bad

CONCLUSION

The development of higher-order thinking skills questions uses the research and development (R & D) method developed by Borg & Gall with a ten-step process. This study aims

to develop higher-order thinking skills questions to increase the cognitive level of students. The products produced in this study are 10 multiple choice questions based on valid and reliable higher-order thinking skills and have met the question criteria in terms of material, construction, and language. The questions also have good quality when viewed from the distinguishing power, the level of difficulty, and the effectiveness of the distractors.

From the results of the research above, the researcher concludes, this validity is used to measure the extent to which the questions measure students' abilities. The results of the validity test of all valid questions, it is known that all items are declared valid with 1 question having high validity, 7 questions having sufficient validity, and 2 questions having low validity. While the results of the reliability analysis using SPSS show the Cronbach's Alpha value of 0.655, then the question is reliable.

The discriminating power test used anatesV4 with the results that all the questions had good discriminating power with a discriminating power index of 44.44-66.67. The results of these scores are included in the criteria for questions with good discriminating power

The results of the analysis show that 9 out of 10 items have a level of difficulty in the "Medium" category and 1 item has a level of difficulty in the "Difficult" category. This analysis tests students' knowledge and problem solving in solving problems.

REFERENCES

Arifin, Z., & Retnawati, H. (2017). Developing an Instrument to Measure Mathematics Higher-order Thinking Skills of 10th Grade Student in Senior High School. *PYTHAGORAS: Jurnal Pendidikan Matematika, 12*(1), 98. https://journal.uny.ac.id/index.php/pythagoras/article/view/14058

Arikunto, S. (2009). *Dasar-Dasar Evaluasi Pendidikan* (9th ed.). PT. Aksara Bumi.

Direktorat Pembina Sekolah Menengah Atas, Direktorat Jenderal Pendidikan Dasar dan Menengah, & Kementerian Pendidikan dan Kebudayaan. (2017). *Modul Penyusunan Soal Higher-order Thinking Skills (HOTS)*.

Iberahim, A. R., Mahamod, Z., & Mohamad, W. M. R. W. (2017). Pembelajaran Abad Ke-21 Dan Pengaruhnya Terhadap Sikap, Motivasi Dan Pencapaian Bahasa Melayu Pelajar Sekolah. *Jurnal Pendidikan Bahasa Melayu, 7*(2), 77–88.

Kadir, A. (2015). Menyusun dan Menganalisis Tes Hasil Belajar. *Jurnal Al-Ta'dib*, *8*(2), 70–81.

Kartowagiran, B. (2012). Penulisan Butir Soal. *Universitas Negeri Yogyakarta*, 1–33.

Lewy, Zulkardi, & Aisyah, N. (2009). Pengembangan Soal Untuk Mengukur Kemampuan Tingkat Berpikir Tingkat Tinggi Pokok Bahasan Brisan dan Deret Bilangan di Kelas IX Akselerasi SMP Xaverius Maria Palembang. *Jurnal Pendidikan Matematika*, *3*(2), 15–28.

Marniwati. (2019). Peningkatan Higher-order Thinking Skills (HOTS) Melalui Strategi Pembelajaran Peningkatan Kemampuan Berpikir (SPPKB) pada Pembelajaran Ekonomi. *Jurnal Pendidikan Tambusai*, *3*(5), 968–975. http://e-journal.uajy.ac.id/14649/1/JURNAL.pdf

National Education Association. (2002). *Preparing 21st-century Students for a Global Society : An Educator's Guide to the "Four Cs."* http://www.nea.org/assets/docs/A-Guide-to-Four-Cs.pdf

OECD. (2015). *Description of the Released Unit from the 2015 PISA Collaborative Problem-Solving Assessment, Collaborative Problem-Solving Skills, and Proficiency Levels*.

https://puspendik.kemendikbud.go.id/pisa/

Oktanisa, L., & Fitrayati, D. (2018). Pengembangan Asesmen Soal Berbasis Higher-order Thinking Skills Pada Mata Pelajaran Ekonomi. *Jurnal Pendidikan Ekonomi*, *6*(3), 355–361.

Schleicher, A. (2018). PISA 2018 : Insights and Interpretations. *Organisation for Economic Cooperation and Development (OECD)*. https://www.oecd.org/pisa/

Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

Suhady, W., Roza, Y., & Maimunah. (2020). Pengembangan Soal untuk Mengukur Higher-order Thinking Skill (HOTS) Siswa. *Jurnal Gantang*, *5*(2), 143–150. https://doi.org/10.31629/jg.v5i2.2518

Supranoto, H. (2018). Pengembangan Soal HOTS Berbasis Permainan Ular Tangga pada Mata Kuliah Telaah Ekonomi SMA. *Jurnal Pendiidikan Ekonomi UM Metro*, *6*(1), 103–110.

Syutharidho, & M, R. R. (2015). Pengembangan Soal Berpikir Kritis untuk Siswa SMP Kelas VIII. *Al-Jabar: Jurnal Pendidikan Matematika*, *6*(2), 219–227.

Trilling, B., & Fadel, C. (2009). 21st-century Skills Learning for Life in Our Times. *Journal of Sustainable Development Education and Research*, *2*(1), 243.

Uno, H. B., & Koni, S. (2013). *Assessent Pembelajaran* (D. Ispurwanti (ed.); 3rd ed.). Bumi Aksara.

Wijaya, E. Y., Sudjimat, D. A., & Nyoto, A. (2016). Transformasi Pendidikan Abad 21 Sebagai Tuntutan Pengembangan Sumber Daya Manusia di Era Global. *Prosiding Seminar Nasional Pendidikan Matematika 2016 ~ Universitas Kanjuruhan Malang*, *1*, 263–278. http://repository.unikama.ac.id/840/32/263-278 Transformasi Pendidikan Abad 21 Sebagai Tuntutan Pengembangan Sumber Daya Manusia di Era Global .pdf. diakses pada; hari/tgl; sabtu, 3 November 2018. jam; 00:26, wib.