The development of higher order thinking skills (HOTS) questions for static fluid concept

M Rizki¹, Sulastri² and Mursal³
¹Departement of Science Education, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia
²Department of Chemistry Education, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia
³Department of Physics Education, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia

E-mail: sulasstri@fkip.unsyiah.ac.id

Abstract. The achievements of Indonesian students at TIMMS are still categorized low. One of the reasons is that students are poorly trained in solving HOTS problems. The observations showed that questions used in the 2017/2018 odd semester exam in high schools in Banda Aceh were still categorized as LOTS, so giving HOTS questions has become a necessity. The purpose of this study was to develop and analyze HOTS questions on static fluid materials. The method used was research and development (R&D) with the ADDIE model. The number of samples was 66 students who were selected using a random sampling technique. The data were collected through questionnaires and written tests. The average percentage of expert validation was 97.98% which indicated a very decent category. The quality of the HOTS question test obtained 16 (20%) valid items and 4 (20%) invalid items. The reliability was 0.878 which indicated a very high category. The test item difficulty index consisted of 6 difficult items, 9 moderate items and 5 easy items. The confounders as much as 92.25% worked well and 8.75% did not work. It can be concluded that the questions from the results of this development can be used to train students HOTS.

1. Introduction

The principles and standards of assessment emphasize the idea of improving student learning and also served as a tool in decision making [1]. Assessment is an important element in educational success [2]. Regarding the issue of the development of education in Indonesia, the 2013 curriculum was designed with various improvements, one of which was adopting an international standard assessment [3]. Referring to this, the assessment in Indonesia is directed at a model of high-level thinking skills (HOTS), in order to produce human resources who have the ability to think critically, creatively, and innovatively, solve problems, make decisions, adapt to the environment and information technology, and have a strong and positive character [4,5].

The fact shows that the quality of science learning in several countries in the world was still relatively low [6]. This can be seen from the position of Indonesia which has not been encouraging in PISA. The results of PISA evaluation of the scientific abilities of Indonesian students in 2012 and 2015 show an average score of 375 and 403, respectively. There is a significant change in the scores...
obtained. However, if it is viewed from all the countries surveyed, Indonesia is still ranked 62 out of 70 countries participating [7,8].

The achievement of Indonesian students at TIMSS is also still categorized low. In 2011 and 2015, the learning achievement scores of Indonesian students were 386 and 397 (scale 0 to 800), respectively, with an average score of 500 [9]. The low achievement at TIMSS is caused by several factors. One of them is Indonesian students are less trained in solving contextual problems that requires reasoning, argumentation and creativity, as the characteristics in TIMSS [10].

The difficulty level of the test questions contained in the National Examination and College Entrance Examination requires students not only to be able to remember concepts and formulas but also to be able to analyze questions that require HOTS thinking. However, in reality, the assessment process in schools, especially in evaluation activities, still provide a lot of LOTS test questions [10]. The problem is that the test questions used in the school assessment still tend to be oriented towards measuring aspects of memory, while the test questions to train student’s high order thinking skills are not sufficiently available [11].

The preliminary research conducted by the researcher found that from all the test questions used in the odd semester exam in 2017/2018 on physics subject at Senior High School 3 Banda Aceh there were level C2 (14.30%), C3 (42.85%), and C4 (42.85%), and there were no questions of level C5 and C6 used in the test. The results of the question quality analysis in 614 high schools in Indonesia obtained only 7.85% of the forms of tests made by teachers of level C4, C5 and C6 in the semester final examinations [12].

The physics lesson at the high school level is one of the subjects that require students to be able to develop reasoning skills in thinking inductive and deductive analysis. Thus, HOTS ability is a priority in studying physics so that students are able to compete in the era of globalization [13]. Static fluid is one of the physics materials that must be studied by high school students of class XI. Most of the material concepts are applied in everyday life, which requires students to think, find, and solve problems based on the relevance of the theories and concepts that they have learned in school [14].

High-level thinking skills are the widespread used skills to find new challenges, which require someone to apply new information or knowledge that he has and manipulate information to reach possible answers in new situations [15,16,17]. High-level thinking is understood as the top end of Bloom’s cognitive taxonomy that aims to enable students to apply knowledge and skills to new contexts, such as applying concepts that have not been taught before [18,19]. Referring to Bloom’s revised taxonomy cognitive assessment, thinking skills are divided into two, namely LOTS and HOTS. The ability of LOTS involves the ability to remember (C1), understand (C2), and apply (C3), while HOTS involves analysis of synthesis (C4), evaluating (C5), and creating or creativity (C6) [20]. The description and keywords of HOTS thinking skills can be seen in Table 1.

| Indicators | Keywords | Category |
|------------|----------|----------|
| Remembering | Can students say and mention definitions, imitate speech, state structure, remember information? repeat, and state Group, describe, explain identification, place, report, explain, and translate. | Lower Order Thinking Skill (LOTS) |
| Understanding | Can students explain the concepts, principles, laws or procedures? | |
| Applying: | Can students apply | Choose, demonstrate, acting, using, illustrate, interpret, schedule, sketch, solve problems, and write. |
| Analyzing: | Can students sort out parts based on differences and | Review, compare, contrast, differentiate, do discrimination, separate, test, do experiments, and asking. |
|            | based on differences and | |

Table 1. Bloom revision taxonomy description and keywords
similarities?

Evaluating Can students declare good or bad against a particular phenomenon or object?

Creating: Can students create an object or view?

Referring to the explanation above, a change in the learning system was needed, especially in the evaluation process and activities as demanded in the 2013 curriculum. The obstacles that existed were the lack of availability of questions that could train students for HOTS. Therefore, providing HOTS questions has become a necessity. In this case the researchers conducted research related to the development of HOTS questions on static fluid material for high school students.

2. Research method
This type of research is Research and Development (R&D) with the ADDIE model which has five stages, namely Analysis, Design, Development, Implementation and Evaluation.

2.1. Analysis
Researchers analyzed the high school physics exam at odd semester 2017/2018 to get information about the cognitive levels found in the questions. The analysis results obtained 14.30% of level C2, 42.85% of level C3, and 42.85% of level C4, and there were no questions of level C5 and C6. This finding was the basis that the questions used to assess is the students were still LOTS. It did not meet the evaluation standards expected in the 2013 curriculum.

2.2. Design
In this stage, the researcher determined the structure of the development of HOTS questions by (1) adjusting the development of HOTS questions with competency standards and basic competencies (2) determining the format in accordance with the HOTS test guide development module (3) conducting the initial design towards the development of HOTS questions (4) making a development HOTS grid test, and (5) preparing expert assessment instruments on HOTS questions.

2.3. Development
Some activities carried out at this stage of development were:

- Composing a grid of HOTS based on the formulation of indicators of static fluid material and HOTS indicators.
- Assembling HOTS items based on the grid, material indicators and HOTS indicators following the previous points using attractive and contextual stimuli. This activity produced 20 multiple choice HOTS question items, each of which was equipped with answers. The narratives about HOTS were presented using interesting and contextual stimuli. These stimuli were in the form of narration or visualization of an event that occurs in everyday life by presenting complex problems to provoke students to think HOTS in solving them [10].
- Checking the qualitative validity of HOTS questions that was conducted by three validators on the construction, material and language of questions using a questionnaire.
- Correcting HOTS items based on suggestions from the validator team. The results of validation by the validator were used to determine the test questions that had been developed whether they still needed improvement or were already feasible to continue to the next stage.

2.4. Implementation
HOTS questions that had been through the validation process were then implemented with a limited trial technique conducted at three state high schools in Banda Aceh. The school and sample selection
were done by a random sampling technique. The total number of samples used in this study was 66 students. The limited test function was to analyze questions quantitatively which included difficulty index, validity of power differences, and reliability of HOTS questions.

2.5. Evaluation
The determination of the feasibility of HOTS questions that had been developed was by using the Excel Proanayltes version 6 [21]. The parameters analyzed after the trial were limited to the power difference, the level of difficulty, validity, reliability and deceptive functions. Furthermore, after going through this analysis activity, the researchers drew the overall conclusion regarding the feasibility of the product about HOTS questions of static fluid material.

3. Results and Discussion
In this section, we present results and discuss them as follows. First, we describe teachers’ internet accessibility at homes and schools. Second, we discuss about teachers’ acceptance to the online professional development course according to their internet accessibility.

In the first stage of this development research, it was produced 20 items of multiple choice of HOTS type called draft 1. The questions contained interesting stimuli and contexts. The stimulus function was to direct students to have the ability to transfer the concepts and cases displayed in the results of the limited test analysis of HOTS question items can be seen in Table 2 [23].

Some small-scale trials were conducted to see construction clarity, language problems and the materials being tested for six high school students. The results of these trials were called HOTS draft II questions. The importance of carrying out small-scale trials was to ensure the prototype functionality that had been developed before the actual trials were carried out [22]. After improving the results of small-scale trials, it was continued with limited trials. Based on the students' answers, the results of the limited test analysis of HOTS question items can be seen in Table 2 [23].

| Question number | Complexity index | Degree of difference | Validity |
|-----------------|------------------|----------------------|----------|
| P index | Category | D values | Category | Information | rxy | Category | Information |
| 1 | 0.500 | Medium | 0.611 | Good | Accepted | 0.538 | Low | Valid |
| 2 | 0.136 | Difficult | 0.500 | Good | Accepted | 0.703 | High | Valid |
| 3 | 0.258 | Difficult | 0.667 | Good | Accepted | 0.677 | High | Valid |
| 4 | 0.530 | Medium | 0.667 | Good | Accepted | 0.504 | Low enough | Valid |
| 5 | 0.712 | Easy | 0.611 | Good | Accepted | 0.462 | Low enough | Valid |
| 6 | 0.697 | Medium | 0.333 | Enough | Revision | 0.330 | Low | Validation |
| 7 | 0.682 | Medium | 0.611 | Good | Accepted | 0.461 | Low enough | Valid |
| 8 | 0.758 | Medium | 0.056 | Less | Rejected | 0.108 | Very low | Invalid |
| 9 | 0.742 | Easy | 0.167 | Less | Rejected | 0.179 | Very low | Invalid |
| 10 | 0.652 | Medium | 0.444 | Good | Accepted | 0.447 | Low enough | Valid |
| 11 | 0.212 | Difficult | 0.611 | Good | Accepted | 0.645 | High | Valid |
| 12 | 0.652 | Medium | 0.556 | Good | Accepted | 0.406 | Low enough | Valid |
| 13 | 0.152 | Difficult | 0.444 | Good | Accepted | 0.561 | Low enough | Valid |
| 14 | 0.682 | Medium | 0.556 | Good | Accepted | 0.470 | Low enough | Valid |
| 15 | 0.894 | Easy | -0.056 | Less | Rejected | -0.143 | Very low | Invalid |
| 16 | 0.121 | Difficult | 0.433 | Enough | Accepted | 0.527 | Low enough | Valid |
| 17 | 0.803 | Easy | 0.167 | Less | Rejected | 0.149 | Very low | Invalid |
| 18 | 0.424 | Medium | 0.667 | Good | Accepted | 0.530 | Low enough | Valid |
| 19 | 0.242 | Difficult | 0.833 | Very Good | Accepted | 0.802 | Very high | Valid |
| 20 | 0.394 | Medium | 0.611 | Good | Accepted | 0.500 | Low enough | Valid |
The information that can be seen from table 3 include the index of difficulty in the test questions with as many as 6 (30%) items were categorized as difficult, 9 (45%) items were categorized as medium, and 5 (25%) items were categorized easy. The difficulty index of good items was categorized as being moderate, meaning that the question was not too easy and not too difficult [24]. Furthermore, different power analysis tests aimed to find out the circumstances of students who truly understood the concepts that they have learned and who did not. The results of the power test were different, where there were 15 (75%) items were categorized as very good and good with very feasible criteria to be accepted as a test that has a high-power difference. However, there was 1 (5%) item that needed an improvement and was suitable to be used, and 4 (20%) items were categorized as lacking, so these questions were not feasible to use. The test results of the validity of the test questions found 16 (80%) items in the test were categorized as valid, while the invalid ones were 4 (20%) items. In addition, the result of the analysis of the reliability of the test questions was 0.870 which indicated a very high category.

The last stage of the analysis was the quality of the deception, there were 5 test questions that were not functional. These questions were found at numbers 6, 8, 9, 15 and 17. This was because the percentage of deceptive items was not chosen up to 5% by the test participants [25]. The overall percentage was 91.25% of deceptive items as that functioned properly and 8.75% that did not function.

One example of a development result is as follows:

There was an accident that caused a car to fall into the lake. The condition of the car is shown in the picture. Determine how much the hydrostatic force is at the central pressure location on the car door! If it is assumed there is no water entering the car and the driver is safe, is it possible that he can come out by opening the car door? (If your choice is possible or not possible, write down your reasons in the answer sheet column provided). (Gravity acceleration 9.8 m/s²).

a. 89.13 N (getting out is not possible)
b. 89.13 N (getting out is possible)
c. 101.20 N (getting out is not possible)
d. 101.20 N (getting out is possible)
e. 127.04 N (getting out is possible)

4. Conclusions
Based on the results of the research that has been done, it could be concluded from the HOTS multiple choice questions, there were 16 valid test items and 4 invalid test items. The overall percentage was 80% of the results of the development of HOTS questions for static fluid material that could be used to train students HOTS. The researchers recommend that the HOTS results of this development should be carried out in extensive trials and they can be used as an evaluation tool for learning activities.

5. References
[1] Budiman A and Jailani 2015 Proc.of International Conf. On Research, Implementation and Education of mathematics and sciences 2015 (Yogyakarta:State University) 81
[2] Bireunbaum M, DeLuca C, Earl L, Heritage M, Klenowsky V, Looney A, Smith K, Timperley H, Volante L and Wyatt-Smith C 2015 J Policy Futures in Educations 13 114
[3] WarisdonoE 2017 Modul Penyusunan Soal Higher Order Thinking Skill (HOTS). (Jakarta: Direktorat Pembinaan SMA Direktorat Jendral Pendidikan Dasar dan Menengah)
[4] Widihasuti 2015 Pros. Seminar Nasional 2015 Jurusan PTBB FT UNY (Yogyakarta: UNY) 77
[5] Widiawati L, Joyoatmojo S and Sudiyanto 2018 International J of Multicultural and Multireligious Understanding 5 96
[6] Tjalla A 2010 Pros.TemaulmiahNasional Guru (Jakarta: UNJ) 1
[7] OECD 2012 PISA 2012 Result in Focus What 15-year-old Know and What They Can do With What They Know
[8] _______2015 PISA Result in Focus What 15-year-old Know and What They Can do With What They Know
[9] Mullis I V S, Martin M O, Foy P and Hooper M 2015 TIMMS 2015 Internasional Result in Mathematic (Boston Colage: International Studi Center)
[10] Widana W I, Partawa I M Y 2018 International J of Social Sciences and Humanities 2 24
[11] Kusuma M D, Rosidin U, Abdurrahman and Suyatna A 2017 IOSR Journal of Research & Method Hooin Education 7 26
[12] Widana I W 2017 J of Indonesia Student Assesment and Evaluation.3 33
[13] Badan Standar Nasional Pendidikan 2006 Peraturan Menteri Nomor 22 Tahun 2006 Tentang Standar Isi (Jakarta: BSNP)
[14] Sulasih, Suparmi A and Sarwanto 2017 International Conf. on Science and Applied Science 909 012060
[15] Kiltz L 2009 J of Homeland Security and Emergency Management 6 1
[16] Heong Y M, Othman W D, Yunos J, Kiong T T, Hassan R and Mohamad M M 2011 International J of Social and humanity 1 121
[17] Hashim H, Ali M N and Shamsudin M A 2017 International J of Academic Research in Business and Social Sciences 7 1192
[18] Athanassiou N and McNett J 2003 J of Management Education 27 533
[19] Brookhart S M 2010 How to Assess Higher Order Thinking Skills in Your Classroom (Alexandria: ASCD)
[20] Anderson L W and Krathwohl D R 2001 A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objective (New York: Longman)
[21] Khaldun I 2017 Proanaltes V6: Program Analisis Butir Soal dan Penilaian Hasil Belajar (Banda Aceh: Jurusan Pendidikan Kimia FKIP Unsyiah)
[22] Davis F D and Venkatesh V 2004J IEEE Transactions of Engineering Management 51 31
[23] Sukmadinata S N 2010 Metode Penelitian Pendidikan (Bandung: Remaja Rosdakarya)
[24] Yusrizal 2016 Pengukuran & Evaluasi Hasil dan Proses Belajar (Yogyakarta: Pale Media Prima)
[25] Sudijono 2001Pengantar Evaluasi Pendidikan (Jakarta: Raja Grafindo Persada)