Developing high order thinking skills (HOTS) assessment instrument for fluid static at senior high school

V Serevina*, Y P Sari and D Maynastiti

Physics Education Study Program, State University of Jakarta, Pemuda Street No.10, Rawamangun – Jakarta Timur, 13220

* vina.serevina77@gmail.com

Abstract. High Order Thinking Skills (HOTS) assessment instrument involve of C4 (analysis), C5 (evaluation), and C6 (creation). This study aims to develop a valid and reliable HOTS assessment instrument and to describe the quality of the assessment on static fluid at grade XI senior high school. This research was performed in Senior High School 9, Tangerang, Indonesia. This developed instrument was used to measure HOTS of students skill in cognitive aspect. Method of this research is research and development (R&D) using Borg and Gall model. Advantage of this research is suitable for a revised 2013 curriculum. The result of this research, 30 items of this instrument were valid from 35 items by using point biserial validation test and reliable by using KR-21realiability test. So, it can be applied in learning process.

1. Introduction
The main purpose of education is to increase human resources, which is influenced by various factors. One of the factors that influence this success is the ability of teachers to perform and utilize assessment, process evaluation and learning outcomes. Such capability is necessary to know whether or not the learning objectives that have been set in the curriculum have been achieved. In addition, the ability can also be used to improve quality of the learning process that has been done by the teacher. Related to that, the government has given guidance by issuing Permendiknas No. 16, 2007 about Academic Qualification Standard and Teacher Competence subjects stated that the competence of subject teachers among others is developing assessment instrument.

Educational Assessment by Permendiknas Number 20 of 2007 on Education Assessment Standards is the process of collecting and processing information to determine the achievement of learners' learning outcomes [1]. The principles and standard of assessment emphasize two main ideas namely that assessment should improve learners' learning and assessment is a valuable tool for making teaching decisions [2]. Assessment is not just the data collection of learners, but also the processing to obtain a picture of the process and learning outcomes of learners. Assessment is not just giving learners questions and then finished, but the teacher must follow up for the sake of learning. To carry out the assessment, teachers need assessment instruments in the form of good questions to test cognitive, affective, and psychomotor abilities.

Assessment is a very important activity in Physics learning. Assessment can provide constructive feedback for teachers as well as learners. The results of the assessment can also provide motivation to learners to perform better. Even judgments can affect learning behavior because learners tend to
direct their learning activities toward the teacher's assessment estuary. The quality of the assessment instrument has directly influence in the accuracy of the achievement status of the teacher outcomes. Therefore, the position of the assessment instrument of learning outcomes is very strategic in making the decisions of teachers and schools related to the achievement of teacher outcomes which include the ability of HOTS (Higher Order Thinking Skills). There is various type of assessments that teachers can make to measure student achievement in learning. When teachers want to measure the cognitive domain of students, teachers can use assessment in the form of tests. Assessment using tests made in the form of written tests as well as non-written tests. Written tests can be multiple choice, description and true or false. Meanwhile, non-written test can be question and answer, oral and observation test.

The situation in the field showed that so far the teacher's judgEment has not been well-directed to measure students 'cognitive abilities, especially on students' Higher Order Thinking Skills. According to Vina Serevina (2016) The success of learning evaluation activities depends on the test instruments used. Test instruments have an important role in measuring student outcomes [3]. So, it is required a standard test instrument and only some teachers who use the correct written test assessments are appropriate in measuring indicators of learning achievement and the cognitive domain of students. The result of observation showed that in Figure 1.

![The Result of Demanding Analysis](image)

**Figure 1. Diagram of Analysis of HOTS Assessment Instrument**

The results of the demanding analysis indicate that 43% of educators in schools are in need of HOTS (Higher Order Thinking Skills), 50% say necessary and 7% say less needless to understand HOTS rating system. It is necessary to prepare a written test assessment that is capable of measuring the ability of HOTS of students and student achievement IN learning indicators. Therefore, it will be conducted research with the title "Development of Higher Order Thinking Skills (HOTS) Assessment Instrument on Static Fluid Material".

Based on the background of the problems presented, the researcher focuses on the development of HOTS assessment instrument on Static Fluid Material of Senior High School 9 at grade 11. Based on the 2013 curriculum of basic competency 3.3 Applying static fluid laws in daily life, basic competency 4.3 Plan and conduct experiments that utilize the properties of static fluids, along with the presentation of results and their physical meanings.

2. Literature Review

2.1 Instrument Development Concepts

According to Borg and Gall research development has steps in its application, namely: 1) Research and information collecting; included in this step include literature studies related to the issues studied, and preparation for formulating research frameworks; 2) Planning; including in this step formulating skills and expertise related to the problem, determining the objectives to be achieved at each stage, and if possible or necessary to conduct a feasibility study on a limited basis; 3) Develop preliminary form of product, which is developing the initial form of the product to be produced. Included in this step is the preparation of supporting components, preparing guidelines and manuals and evaluating the
feasibility of supporting equipment; 4) Preliminary field testing, conducting initial field trials on a limited scale, involving the subject from 6 to 12 subjects. In this step, data collection and analysis can be done by interview, observation or questionnaire; 5) Main product revision, make improvements to the initial product produced based on initial test results. This improvement is most likely to be done more than once, in accordance with the results shown in the limited trials, in order to obtain a draft of the main product (model) ready for a wider trial; 6) Main field testing, the main test involving all students. 7) Operational product revision, which is to make improvement of the test results more widely, so that the developed product is a design operational model that is ready to be validated; 8) Operational field testing, which is a validation test of the operational model that has been produced; 9) Final product revision, making final improvements to the model developed to produce the final product; 10) Dissemination and implementation, which is the step of disseminating product/develop model [4].

2.2 Assessment Instruments

The form of assessment instrument consists of test and non-test instruments. Form of assessment instruments developed in this study using multiple choice test instruments. Multiple options can be used to measure HOTS or high-level thinking skills, according to Brookhart [5], Nitko & Brookhart [6] and Kubiszyn & Borich [7].

Three item formats or tasks in high-order thinking skills are useful in measuring higher order skills: (a) selection, which includes multiple choice, matching and ranking sequence items; (b) generation, which includes tasks and tasks of short answers, essays, and performance; and (c) explanations, which involve giving reasons for election or generational responses [8].

HOTS test item organized by HOTS indicators and Basic Competency (KD) indicators. HOTS indicator synthesized from indicators of critical and creative thinking by Nitko & Brookhart [11], Arends & Kilcher (2010), Presseisen [9], Szetela [10], Krulik & Rudnick [11], O’Daffer & Thornquist [12], Maite & Laura [13] and Perkins [14]. The indicator is meant, among others: (1) identify and associate the relevant information from a situation/problem, (2) make the right conclusions based on the information of a situation/problem, (3) find the consistency/inconsistency in an operations/products, (4) assess an operation/relevant products based on criteria/standards, (5) blends the ideas/strategies to solve a problem, (6) using the ideas/the right strategies to solve a problem, (7) develop or create new alternative in resolving a problem.

2.3 HOTS (Higher Order Thinking Skills)

According to Conklin HOTS characteristics are as follows: “characteristics of higher-order thinking skills: higher order thinking skills encompass both critical thinking and creative thinking” [15]. There are four components of HOTS (higher order thinking), as follows [16]:

![Figure 2. Component Higher Order Thinking Skills](image)

Crowl defines critical thinking as part of the process of evaluating evidence gathered in problem-solving or outcomes generated by creative thinking. The self-correcting nature of self-correction is called "metacognition." Metacognition includes awareness of one thinking process, self-monitoring, and the application of known heuristics and thought steps. One’s success with metacognition depends,
in part, on the belief in one's ability to be smarter and the beliefs of others, such as the teacher, in one's ability.

According to Wang and Wang (2011), there are three main components in HOTS, i.e. critical thinking skills, design thinking skills and system thinking skills, while Miri et al. (2007), states that HOTS consists of three components, namely critical thinking skills, systematic thinking skills and creative thinking skills. Furthermore, according to Yee Mey Hong et al. (2011), critical thinking skill and creative thinking skill are two important indicators of HOTS. Thus, there are at least two indicators in HOTS, so that measurement of the students’ HOTS can be conducted by observing their critical and creative thinking skills. HOTS is a latent variable that can’t be measured directly. In order to measure the characteristics of latent variables, according to Naga (2012), manifest variables can be used to be measured latent variables. Measurement of manifest variables requires a standardized instrument. The problem now is how to provide an instrument to allow teachers to measure students’ HOTS.

Characteristics of HOTS expressed Resnick (1987, p.3) which are non-algorithmic, complex, multiple solutions (many solutions), involves a variety of decision making and interpretation, application of multiple criteria (many criteria), and effortful (requires a lot of effort). Conklin states the following characteristics of HOTS: “characteristics of higher-order thinking skills: higher order thinking skills encompass both critical thinking and creative thinking”[17]. Critical and creative thinking are two very basic human capabilities because both can encourage someone to always look at every problem faced critically and trying to find the answer creatively in order to obtain a new thing better and beneficial for life.

3. Research Methods
This research is a development research. The product developed based on a HOTS (Higher Order Thinking Skills) assessment instrument in the form of HOTS test of multiple choice and HOTS test. To get the development prototype, this research is an adaptation of Borg & Gall development model. From 10 steps the development of Borg & Gall model is adapted into seven development steps: 1) research and information gathering, 2) planning 3) initial product development, 4) limited testing, 5) initial product revision, 6) field trial and 7) revision of the final product. Research and information gathering are conducted for the study of concepts based on relevant theoretical studies. Validation of assessment instruments is conducted to evaluate the validity and evaluation instruments in the form of items on HOTS test. Validation is done at the initial product development stage by three physics education experts. An empirical test of HOTS items was conducted using a limited trial and field trials. Trials were conducted at Senior High School 9 Tangerang. Analysis of trial data using classic test theory parameters to determine the quality of HOTS test questions empirically as a basis for revision and assembly of HOTS tests.

| No | Aspects of HOTS Ability | HOTS Indicator | Purpose of Learning | Item Number |
|----|-------------------------|----------------|---------------------|-------------|
| 1. | Analyze | Analyze information and structure information into its smaller sections to recognize patterns or relationships | Determine the density of a liquid by using the corresponding vessel principle | 1, 20, 21 |
| 2. | Evaluate | Load hypothesis, criticize, and test | Compare the lifestyle values with various physical phenomena | 2, 19, 22 |
| 3. | Evaluate | Accept or reject a statement based on established criteria | Predict an event based on the legal application of Archimedes | 3, 18, 23 |
4. Analyze
Being able to recognize and distinguish factors and causes from a complicated scenario
Solve the phenomenon of physics in accordance to the concept of:

4. Hydrostatic pressure
4, 17, 24

5. Analyze
Being able to recognize and distinguish factors and causes from a complicated scenario
Solve the phenomenon of physics in accordance to the concept of:

5. Pascal law
5, 16, 25

6. Analyze
Being able to recognize and distinguish factors and causes from a complicated scenario
Solve the phenomenon of physics in accordance to the concept of:

6. Law of Archimedes
6, 15, 26

7. Analyze
Being able to recognize and distinguish factors and causes from a complicated scenario
Solve the phenomenon of physics in accordance to the concept of:

7. Viscosity law
7, 14, 27

8. Analyze
Being able to recognize and distinguish factors and causes from a complicated scenario
Solve physics phenomena according to the concept of:

8. Pressure
8, 13, 28

9. Analyze
Being able to recognize and distinguish factors and causes from a complicated scenario
Solve the physics phenomenon according to the concept of:

9. Pressure on the hydraulic pump
9, 12

10. Analyze
Being able to recognize and distinguish factors and causes from a complicated scenario
Solve the physical phenomena according to the concept of:

10. Surface tension
10, 11

11. Create
Able to create a scientific work and experiment
Make scientific work, create your own experiments

29, 30

4. Results and Discussion

4.1 Results
The result of the development in this research is the HOTS (Higher Order Thinking Skills) assessment instrument in the form of a matter of multiple choice test of Static Fluid Material of grade 11 at senior high school which is valid and reliable. The developed assessment instrument has passed two stages of assessment. Assessment of the first stage is conducted to assess the validity and evaluation instruments performed by Physics education experts. Assessment of the second phase of field tests conducted HOTS. The processes undertaken in this development include the preparation of HOTS test products.

The processes undertaken in this development include the preparation of HOTS test products (Higher Order Thinking Skills). The HOTS test was designed to be assessed by expert validators, revised for the initial HOTS test items ready for use as a limited trial material. The results of the trial are limited, as a revision material to then become the main product of the HOTS test which ready for use as field test materials.

4.2 Discussion
Expert validation is performed to view the contents of the original product. This validation aims to obtain input, improvement suggestions, and simultaneous assessment of the initial product prior to a limited trial. Validation activity is done by giving the initial product script that is in the form of lattice problem and test of HOTS (Higher Order Thinking Skills) and validation sheet to three expert validator.
Table 4.1 Suggestion from Validators of HOTS Rating Instrument

| No | Validator | Input Suggestion |
|----|-----------|------------------|
| 1  | Expert 1  | Problem is good enough, should be accompanied by the provision of reasons. |
| 2  | Expert 2  | It's good about HOTS (Higher Order Thinking Skills) in essays |
| 3  | Expert 3  | Some questions are based on HOTS (Higher Order Thinking Skills), only the language should be easier to understand. |
| 4  | Expert 4  | Problem No. 27,8,9 is well revised. |
| 5  | Expert 5  | Some problems have not HOTS (Higher Order Thinking Skills), should be revised. |

The validation process is done by filling out the questionnaire by the validator of the HOTS (Higher Order Thinking Skills) rating instrument on the Static Fluid material. Expert validation results are shown in the graph below:

![Graph of Expert Validation Results](image)

**Figure 4.1** Graph of Expert Validation Results

From Figure 4.1 it can be seen that for the material conformity aspect with the purpose of measurement, the suitability of the instrument item with the indicator, and the language used in the HOTS (Higher Order Thinking Skills) assessment instrument on the Static Fluid material obtained the average value of the validity of 81.2% small-scale tests were conducted after the instrument was revised.

Table 4.2 Test Results Validity Problem

| Number of Questions | r_{table} | r_{table} | Information |
|---------------------|-----------|-----------|-------------|
| 1                   | 1.274     | 0.683     | Valid       |
| 2                   | 1.66      | 0.683     | Valid       |
| 3                   | 2.092     | 0.683     | Valid       |
| 4                   | 1.301     | 0.683     | Valid       |
| 5                   | 1.53      | 0.683     | Valid       |
| 6                   | -1.13     | 0.683     | Drop        |
| 7                   | -0.25     | 0.683     | Drop        |
The result of validity test of instrument there were 5 items that drop, so became 30 items were valid, then this instrument is tested reliability. This reliability illustrates the consistency of an assessment instrument, meaning that in whatever dimension the measured results will remain the same.

| Number of Questions | r_calculated | r_table | Information |
|---------------------|--------------|---------|-------------|
| 8                   | 2.16         | 0.683   | Valid       |
| 9                   | 0.276        | 0.683   | Valid       |
| 10                  | 2.834        | 0.683   | Valid       |
| 11                  | 1.09         | 0.683   | Valid       |
| 12                  | 2.945        | 0.683   | Valid       |
| 13                  | 2.131        | 0.683   | Valid       |
| 14                  | 3.807        | 0.683   | Valid       |
| 15                  | 2.067        | 0.683   | Valid       |
| 16                  | 2.211        | 0.683   | Valid       |
| 17                  | 3.014        | 0.683   | Valid       |
| 18                  | 2.422        | 0.683   | Valid       |
| 19                  | 1.311        | 0.683   | Valid       |
| 20                  | -0.29        | 0.683   | Drop        |
| 21                  | 3.67         | 0.683   | Valid       |
| 22                  | 0.563        | 0.683   | Valid       |
| 23                  | 1.165        | 0.683   | Valid       |
| 24                  | 2.263        | 0.683   | Valid       |
| 25                  | 2.134        | 0.683   | Valid       |
| 26                  | 3.118        | 0.683   | Valid       |
| 27                  | 2.067        | 0.683   | Valid       |
| 28                  | 1.067        | 0.683   | Valid       |
| 29                  | 1.623        | 0.683   | Valid       |
| 30                  | 2.091        | 0.683   | Valid       |
| 31                  | 1.303        | 0.683   | Drop        |
| 32                  | 3.033        | 0.683   | Valid       |
| 33                  | 1.535        | 0.683   | Valid       |
| 34                  | 1.754        | 0.683   | Valid       |
| 35                  | 1.198        | 0.683   | Drop        |

The result of validity test of instrument there were 5 items that drop, so became 30 items were valid, then this instrument is tested reliability. This reliability illustrates the consistency of an assessment instrument, meaning that in whatever dimension the measured results will remain the same.

Table 4.3 Level of Problem HOTS Tests (Higher Order Thinking Skills)

| Category                  | Item Number                  | Total | Percentage % |
|---------------------------|------------------------------|-------|--------------|
| TK < 0.25 (hard)          | 2,6,17,20,21,22,32,33        | 8     | 22.85%       |
| 0.25 ≤TK≤0.80 (medium)    | 1,3,4,5,8,9,10,11,12,13,14,15,16,18,19,23,24,25,26,27,28,29,30,31,34,35 | 26    | 74.28%       |
| TK>0.80 (easy)            | 7                            | 1     | 2.87%        |
Based on Table 4.3 it can be seen that the level of difficulty range in the medium category of 26 items (74.28%). Calculate the level of difficulty of this problem to differentiate between highly-capable learners with low-ability learners.

The grains distinguishing power is known by looking at the correlation coefficient. In general, level of difficulty multiple choice items can be seen in Table 4.7 below.

| Category                     | Item Number                                      | Total | Percentage % |
|------------------------------|--------------------------------------------------|-------|--------------|
| DP < 0.40 (good)             | 1,2,3,4,8,10,11,12,13,14,15,16                   | 12    | 34.30%       |
| 0.30 ≤ DP ≤ 0.39 (received without revision) | 5,9,17,18,23,32,33,34,26,27,2819,21,22,24,25,29,30 | 18    | 51.42%       |
| 0.20 ≤ DP ≤ 0.29 (problem fixed) | 6,7,20,31,35                                   | 5     | 14.28%       |
| DP > 0.19 (replaced / completely revised) |                                       |       |              |

Based on Table 4.4 it can be seen that the level of difficulty ranges in categories received without revision as much as 18 items (51.42%). Based on the analysis of the problem declared eligible to use. Nevertheless, there are some that need to be fixed. A product revision is made to obtain a final product that meets both valid and reliable criteria. Revisions are made on the basis of assessment and analysis of assessment instruments at each stage of the product trial. The product revisions in this study consisted of revisions of the product of validation, revision of the product of a limited trial, and revision of the product of field trials.

The final product of this development research is the HOTS (Higher Order Thinking Skills) physics assessment instrument in the form of HOTS test device. Based on the results of expert validation, limited trials, field trials, and improvements, and data analysis performed it can be seen that the HOTS developed test device has satisfied with the valid and reliable criteria also the quality of the item is good. The quality of assessment instruments in the form of HOTS test questions based on the results of the analysis of the main product item about the HOTS test is to analyze all items based on empirical data. This assessment instrument has a reliability coefficient of 0.964 (very high) on multiple choice questions. The multiple-choice question has an average of 0.580 (medium) difficulty, average distinguishing power of 0.330 (good). The result of this research showed that HOTS (Higher Order Thinking Skills) assessment instrument consist of 30 items of multiple choice on static fluid material.

5. Conclusions
Based on the result and discussion, it can be concluded that a developed HOTS (Higher Order Thinking Skills) assessment instrument with 30 items of this instrument were valid from 35 items by using point biserial validation test and reliable by using KR-21 reliability test. So, it can be applied in learning process.

References
[1] Depdiknas, Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 20 Tahun 2007, tentang Standar Penilaian Pendidikan. Jakarta, 2007.
[2] J. A. Van de Walle, Elementary and middle school mathematics: teaching developmentally, (6th ed.). United States of America: Pearson Education, Inc., 2007.
[3] C. D. Vina Serevina, “Rancangan Tes dan Evaluasi Fisika yang Informatif dan Komunikatif pada Materi Kinematika Gerak Lurus,” JPPPF-Jurnal Penelit. Pengemb. Pendidik. Fis., vol. 2, 2016.
[4] M. . Borg, W. R. & Gall, Educational researcher: An introduction, (4th ed.). New York: Longman, 1989.
[5] S. M. Brookhart, *How to assess higher order thinking skills in your classroom*. Virginia USA: SCD Alexandria, 2010.

[6] S. M. Nitko, A. J., & Brookhart, *Educational assessment of student*. (6th ed.). Boston: Pearson Education, 2011.

[7] G. D. Kubiszyn, T. & Borich, *Educational testing & measurement. Classroom application and practice*. (10th ed.). New York: John Wiley & Sons, Inc., 2013.

[8] F. R. FJ King, Ludwika Goodson, “Higher Order Thinking Skills,” *A Publ. Educ. Serv. Progr.*, p. 3.

[9] B. Z. Presseisen, *Thinking skills: meanings and models*. Alexandria, Virginia: ASCD, 1985.

[10] W. Szetela, *Facilitating communication for assessing critical thinking in problem solving*. Reston, VA: The National Council of Teachers of Mathematics, Inc., 1993.

[11] J. A. Krulik, S., & Rudnick, *Innovative tasks to improve critical and creative thinking skills*. 1999.

[12] B. A. O’Daffer, P. G., & Thornquist, *Critical thinking, mathematical reasoning, and proof*. New York: Maxwell Macmillan International, 1993.

[13] B. Maite, G & Laura, “Effect of a play program on creative thinking of preschool children,” *J. Psychol.*, vol. 14, p. 608, 2011.

[14] D. N. Perkins, *What Creative Thinking Is*. Alexandria, Virginia: ASCD, 1985.

[15] W. Conklin, *Higher-order thinking skills to develop 21st century learners*. Huntington Beach: Shell Educational Publishing, 2012.

[16] G. Schraw, *Assessment of Higher Order Thinking Skills*. The University of Texas at Austin, 2011.

[17] S. Arikunto, *Dasar-Dasar Evaluasi Pendidikan (Edit Revisi)*. Jakarta: Bumi Aksara, 2009.