Development and validation of students’ achievement, ability to ask and inductive thinking instruments in the static fluid course

A Halim1,4,*, D S Ayunda2, Susanna1, Mustafa1, M Syukri1, and Irwandi3,4

1Department of Physics Education, Training Teacher and Education Faculty, Universitas Syiah Kuala, Banda Aceh, Indonesia 23111
2Department of Science Education, Graduate School, Universitas Syiah Kuala, Banda Aceh, Indonesia 23111
3Department of Physics, Faculty of Science and Mathematics, Universitas Syiah Kuala, Banda Aceh, Indonesia 23111
4The Centre of STEM Study, Universitas Syiah Kuala, Banda Aceh, Indonesia 23111

* abdul.halim@unsyiah.ac.id

Abstract. The purpose of this article is to discuss the results of developing several test and non-test instruments to be used as a measurement of learning outcomes, questioning skills and students’ inductive thinking skills. Development using the ADDIE model. Based on expert assessments and trials on a limited sample, 20 items were obtained for the learning outcomes instrument, 12 items for the non-test question ability and 10 items for the non-test instrument inductive thinking ability. The results of the analysis of data for the Learning Outcomes Test of Static Fluid (LOTSF), the Questioning Ability Observation Sheet instrument (QAOS), and the Inductive Thinking Ability Observation Sheet (ITAOS) instrument obtained a validity of 0.6 and a reliability value of 0.75. Inference from this study to lecturers or teachers who teach Static Fluid concepts can follow the test development procedure offered in this article or use instruments produced from this study.

1. Introduction

Government Act no. 21 of 2013 concerning the national education system said that the aim of higher education is to prepare generations who have knowledge and master technology, have high dedication and can overcome problems globally [1,2]. To achieve the results expected by the law, every level of education must have good human resources, supportive references and extensive educational knowledge, and have an appropriate management or evaluation system [3,4,5]. A good evaluation system covers all aspects that need to be present and is mastered by each student, including cognitive, psychomotor and affective aspects [6,7,8].

The development of cognitive domain measurement instruments has been carried out with various forms of methods in the world of education with the aim of getting the best and valid instruments to capture the ability or have students’ knowledge of the concepts that have been taught. Among the development of cognitive test instruments with three-phase techniques; draft development, limited sample trials and extensive sample trials [9], using five main components; academic skills, study,
academic motivation, intrapersonal skills, and self-concept [10], using guidelines from the Standards for Educational and Psychological Testing [11], using learning-oriented scales [12], and using Physics-Scientific Literacy Achievement techniques Test [13].

In addition, the development of instruments to measure "the ability to ask students" has been studied and developed with different models or methods, including through questioning training for teachers that have an impact on students [14,15], by utilizing questioning skills for concept change [16,17], through synchronization and online system utilization [18,19], utilization of productive thinking simulation [20], through the use of "question boards" [21], and through online discussion forums [22,23].

Furthermore, the study of the ability to think inductively has also been developed through various methods of instruments, including; measurement of inductive reasoning through specific domain pursuit strategies [24], with the teaching system of inductive thinking to high-ability students [25,26], the inductive thinking model is measured from the achievement of student learning outcomes [27,28], measured from aspects of reasoned skills and data representation [29], the effectiveness of inductive thinking is measured from the aspect of achieving student learning outcomes [30], using the "Door and Key" test [31], and the impact of using a calculator on inductive thinking [32].

Based on some previous research, studies are rarely focused on developing instruments to measure students' questioning abilities as well as instruments to measure inductive thinking skills systematically. Most studies only focus on the application of certain models to increase the ability to ask questions and what is measured is the student's final learning ability, as well as inductive thinking skills. Therefore, through this study, the development of both instruments has been structured, starting from theoretical and operational definitions to trials on limited samples.

2. Method

2.1 research approach

The approach used in this research is a quantitative research approach with the Research and Development (R&D) method and the research design is based on several previous studies [38,39,40], as shown in Figure 1.

![Figure 1. Summary of the development of instruments in general use [2,33,34,35]](image_url)

The initial draft of the instrument of the three variables above is stated in the form of the Learning Outcomes Test of the Static Fluid material (LOTSF), the ability to ask in the form of the Questioning Ability Observation Sheet instrument (QAOS), and the students' inductive thinking ability measured
using the instrument of the Inductive Thinking Ability Observation Sheet (ITAOS). The draft of the three instruments is shown in table 1.

| Types      | Name of Instruments                                      | Used for                                           |
|------------|-----------------------------------------------------------|----------------------------------------------------|
| Test       | Learning Outcomes Test of the Static Fluid material (LOTSF) | Measuring the level of mastery of the concept of static fluid |
| Non-Test   | Questioning Ability Observation Sheet (QAOS)              | Measuring “the ability to ask students”             |
| Non-Test   | Inductive Thinking Ability Observation Sheet (ITAOS)       | Measuring students’ inductive thinking ability      |

2.2 validation of instrument

Ratification of instruments in the form of tests or non-tests begins with the validity of the evaluation experts and also the content experts. The next stage of validation is a trial on a limited sample (28 students of class X Madrasah Aliyah Private Darul Ihsan in 2019/2020) using the pre-experimental method and selecting the design of one group prestest-posttest [17].

The instrument trial group was taught using the brainstorming method in three meetings or 3 hours face to face (X). Before the treatment is conducted, respondents are given a pre-test (Q1, Q2, Q3) and after the treatment is given a post-test (Q4, Q5, Q6) for three times the implementation of learning using the brainstorming method. The pre-test consists of Q1 is an initial understanding test about static fluid concepts, Q2 is the initial ability to ask questions and Q3 is the initial critical thinking ability of students. The same is true for the post-test consisting of Q4 is a final understanding test regarding the concept of static fluid, Q5 is the final ability to ask questions and Q6 is the final ability of students’ critical thinking

2.3 learning outcomes test of the static fluid material (LOTSF)

The instrument for measuring student learning outcomes on Static Fluid material uses the LOTSF instrument (Learning Outcomes Test of Static Fluid). This LOTSF instrument initially consisted of 30 items developed by researchers by adopting some of the results of previous studies [36,37,38]. LOTSF instrument characteristics consist of 5 items related to force and pressure definitions, 5 items related to force and pressure equations, 10 items related to calculation problems about force and pressure, and 10 items related to contextual examples of static fluid pressure and force. After being given to experts on Static Fluid material, the LOTSF instrument was tested on a limited sample to obtain data on difficulty index, different power index, validity and reliability index [39,40,41].

2.4 analysis of LOTSF test items

The research data were analyzed using excel software version window 10.00 using the percentage equation and also the normalized N-Gain formula [14,39]. Data quantization of test results with LOTSF instruments were analyzed using different power index formulas, difficulty index of test items and correlation formulas to determine the validity of each item. The difficulty level is calculated using the formula [39]. The P variable is expressed as a difficulty index, B is the number of students who answer the questions correctly, and S is the total number of students taking the test. The interpretation of the results of the analysis with the formula refers to the provisions on the classification of the difficulty index as shown in figure 3 [39]. While the calculation of the different power index test items is used the following equation [39]. With provisions for PA and PB associated with the following formula [39]. Variable D is discriminatory testee (item different power number), PA is the proportion of upper group testees who can answer correctly, PB is the proportion of lower group testees who can answer correctly, BA the number of upper group testees who can answer correctly, BB number of lower group testees which can answer correctly, JA number of testees included in the upper group, B the number of testees included in the lower group. The interpretation of calculation results using the difference power index formula refers to [39].
Furthermore, item analysis or to obtain the validity of the test items uses the product moment correlation formula which is often used to find the relationship between two variables. The validity of an item from a test is the accuracy of the measurement possessed by an item (which is an inseparable part of the test as a totality), in measuring what should be measured through the items [39]. Where \( r_{xy} \) is the correlation coefficient between the variable \( x \) and the variable \( y \), \( \sum xy \) is the number of times \( x \) with \( y \), \( x^2 \) squared of \( x \), \( y^2 \) squared of \( y \). The interpretation of the results of the validity analysis using the equation above [39]. Whereas the N-Gain formula is used to obtain information related to improving learning outcomes after learning with the brainstorming method [42,43]. The results of the analysis of the addition of static fluid concept additions to the N-Gain formula are interpreted by referring [42,43].

2.5 instrument of questioning ability observation sheet (QAOS)

The non-test instrument used to measure students' questioning abilities is in the form of a Student's Questioning Ability Observation Sheet or QAOS. The QAOS instrument developed by researchers consisted of 12 items of student activities that needed to be observed, including 6 items related to mathematical questioning activities, 3 related items how to ask questions about existing information, and 3 items related to asking questions based on information that had been obtained.

| No. | Activity Observed                                                                 | Frequency |
|-----|-----------------------------------------------------------------------------------|-----------|
| 1   | Students ask about events they often encounter in daily life                       | 1 2 3 4   |
| 2   | Students ask about real examples in everyday life                                  |           |
| 3   | Students ask how to apply the material being studied in everyday life              | 1         |
| 4   | Students ask about mathematical equations about the material being studied         |           |
| 5   | Students ask about the object being studied                                       |           |
| 6   | Students ask about real information related to the material being studied         |           |
| 7   | Students ask about the possibilities that occur in everyday life by relating to the material being studied | 1 2 3 4   |
| 8   | Students ask about their predictions about the material being studied by relating it to everyday life |           |
| 9   | Students ask about whether their suspicions about the material being studied can be applied in everyday life | 1 2 3 4   |
| 10  | Students ask about whether the material they are learning contributes to everyday life | 1 2 3 4   |
| 11  | Students ask about the material being studied whether it gives a big advantage in everyday life |           |
| 12  | Students ask about the material being studied not all of them apply in everyday life |           |

The observer records the situation by filling in the observation sheet based on a Likert scale (1 = Never, 2 = Sometimes, 3 = Often, 4 = Very Often). The development of the non-test instrument is based on some of the results of previous studies [44,45,46] and the results of the development of Observation Sheets (QAOS) to measure the ability to ask students after being revised based on input from experts shown in table 2.

2.6 instrument of inductive thinking ability observation sheet (ITAOS)

The non-test instrument used to measure students' inductive thinking abilities is in the form of the Student's Inductive Thinking Ability Observation Sheet or ITAOS. The ITAOS instrument developed by researchers consisted of 10 items of student activities that needed to be observed, including 4 items related to activities on how to obtain information, 2 items related to how to record information, and 3 items related to writing conclusions based on information already obtained. The observer records the state of the respondent by filling out the observation sheet based on a Likert scale (1 = Never, 2 = Sometimes, 3 = Often, 4 = Very Often). The development of the non-test instrument is based on some of the results of previous studies [33,34,35].
3. Results and Discussion

The results of the LOTSF instrument trial to obtain different power indexes, difficulty index and validity are shown in the following figure 2. The trial results show that there are items that need to be revised, there are items that must be discarded due to low validity, different power and difficulty indexes, and items that can continue to be used. Based on consideration of the value of validity that is too low, it cannot distinguish between clever and less clever students and items that are too easy or too difficult, so there are 10 items that must be discarded (bold print in figure 2). Thus only 20 items can be used as a gauge of initial ability or pretest and also to measure ability after the treatment or posttest. The results of item analysis are used to form a graph of ID (different power index) and IK (item difficulty index) as shown in figure 3 [35].

| No. | Item | Validity | Different Power | Difficulty | Ket. |
|-----|------|----------|-----------------|------------|------|
| 1   | Valid| 0.38     |     | 0.31 | Moderate Used |
| 2   | Not Valid| 0.27     | Medium | 0.19 | Bad Deleted |
| 3   | Not Valid| 0.31     | Difficult | 0.19 | Moderate Used |
| 4   | Valid| 0.25     |     | 0.19 | Bad Deleted |
| 5   | Valid| 0.22     |     | 0.19 | Bad Deleted |
| 6   | Valid| 0.22     |     | 0.19 | Bad Deleted |
| 7   | Valid| 0.28     |     | 0.6  | Bad Deleted |
| 8   | Valid| 0.41     |     | 0.6  | Good Used |
| 9   | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 10  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 11  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 12  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 13  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 14  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 15  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 16  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 17  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 18  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 19  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 20  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 21  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 22  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 23  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 24  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 25  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 26  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 27  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 28  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 29  | Valid| 0.22     |     | 0.6  | Bad Deleted |
| 30  | Valid| 0.22     |     | 0.6  | Bad Deleted |

Figure 2. The pilot test results of the multiple-choice test instruments (LOTSF)

Good items are neither too difficult nor too easy, thus the difficulty index of items that are worth using according to table 5 above is between 0.31 and 0.71 and these indices become the lower and upper limits along the vertical axis (IK axis) on the graph IK and ID in figure 3.

While according to the version of the ability of different power, a good item is that it can differentiate answers between students who are smart with answers of students who are low in ability. The different power indices that meet the criteria above are between 0.21 and 1.00. The two indexes are the lower and upper boundaries along the horizontal axis (ID axis) on the IK and ID charts in figure 3. The items in the red box on the IK and ID charts are fit to be used as a test of the ability to master the Static Fluid concept without revision. While items that are outside the red box, but still close to one side of the red box are still suitable for use but need to be revised. Based on the data on the IK and ID graphs in figure 3 and the consideration of the correlation value r between items, 20 items are obtained that are eligible to be used to measure the ability to understand the concept of Static Fluid.

The ability to ask each student is measured by three indicators namely gathering information (factual and procedural), organizing information (objectives and assumptions) and expanding information
(hypotheses and speculations) [22]. The three indicators are scattered in each question item. The average value of the ability to ask students increased significantly at each meeting. It can be understood that the instrument used to measure the ability to ask students (QAOS) has a good reliability index, because in repetition three times the observation gives results that are not significantly different.

Then the average value of the results of observations of students' inductive thinking for the experimental class at the first, second, and third. The results of the analysis show the ability to think inductively measured using ITAOS is not much different between the first, second and third meeting. It can be understood that the instrument used to measure students' inductive thinking skills (ITAOS) has a good reliability index, because in repetition three times the observation gives results that are not significantly different.

The difference in mean values is only due to the effect of learning using the brainstorming method. This brainstorming method is basically aimed at building students' cognitive mentality, so it is very suitable for developing thinking skills, especially inductive thinking. The increase in inductive thinking that occurs in the experimental class proves the effectiveness of the application of brainstorming methods in training these inductive thinking [25,27,33,26].

![Figure 3](image.png)

**Figure 3.** Chart between IK (Difficulty index) and ID (different Power Index) items [35]

4. Conclusion
The development of test and non-test instruments begins with the selection of variables to be measured, theoretic and operational definitions of the selected variables, the determination of indicators in accordance with the previous definitions and the development of instrument items. Theoretical and operational definitions refer to theories that explain the measured variables. The next step is expert validation and trials on a limited sample.

5. References
[1] Depdiknas 2003 *Kebijakan Departemen Pendidikan Nasional* (Jakarta: Biro LN dan Humas)
[2] Yusrizal, Halim A, Daud and Saminan 2020 *J. Phys. Conf. Ser.*, 1460 012112.
[3] Mazlina H, Halim A and Yusrizal 2018 *J. Penelit. Fis. dan Apl.*, 8, 106-111.
[4] Halim A, Lestari D and Mustafa 2019 *J. Phys. Conf. Ser.*, 1280 052060.
[5] Arifullah, Halim A, Syukri M and Nurfadilla E 2020 *J. Phys. Conf. Ser.* 1460 012144.
[6] Yusrizal and Halim A 2009 *J. Pendidik. Malaysia* 34, 33–47.
[7] Halim A, Wahyuni A and Balqis N N 2020 *AIP Conf. Proc.* 050003.
[8] Nurmalia, Halim A and Syahrun N 2020 *J. Phys. Conf. Ser.* 1460 012142.
[9] Rovai A P, Wighting M J, Baker J D and Grooms L D 2009 *the Inter and Higher Edu* 12, 7–13.
[10] DiPerna J C and Elliott S N 1999 *Journal of Psychoeducational Assessment* 17, 207–225.
[11] Cataloglu E 2001 *Development and validation of an achievement test in introductory quantum mechanics: The Quantum Mechanics Visualization Instrument (QMVI).* Dissertation of Doctor of Philosophy (New York: Penn State departments and colleges)
[12] Midgley C, Kaplan A, Middleton M, Maehr M L, Urdan T, Anderman L H and Roeser R 1998 Contemporary Educational Psychology 23, 113–131.
[13] Adeleke A A and Joshua E O 2015 Journal of Education and Practice 6, 28-42.
[14] Mauigoa L and Tekene 2006 New Zealand Journal of Teachers’ Work 3, 12-23.
[15] Mazlina H, Halim A and Yusrizal 2018 J. Penelit. Fis. dan Apl. 8, 106-111.
[16] Yip D Y 2004 Journal of Biological Education 38, 76–83.
[17] Halim A, Yusrizal, Mazlina H, Melvina and Zainaton 2018 J. Phys. Conf. Ser. 1088 012109.
[18] Wang C H 2005 Journal of Computer Assisted Learning 21, 303–313.
[19] Resta N N, Halim A, Mustafa and I. Huda I 2020 J. Phys. Conf. Ser. 1460 012131.
[20] Chin C 2007 Journal of Research in Science Teaching 44, 815–843.
[21] Dixon N 1996 Primary Science Review 44, 8-10.
[22] de Jesus H P and Aurora C M 2012 Reflection Education 8, 57-77.
[23] Halim A, Mustafa, Nurulwati, Soewarno and Nanda N 2018 J. Phys. Conf. Ser. 1120 012030.
[24] Klauer K J 1990 European Journal of Psychology of Education 5, 23-34.
[25] Klauer K J 1992 European Journal of High Ability 3, 164–180 (1992).
[26] Halim A, Yusrizal, Susanna and Tarmizi 2016 Jurnal Pendidikan IPA Indonesia 5, 1-5.
[27] Harneet B 2013 International Journal of Education and Psychological Research 2, 49-59.
[28] Halim A, Ngudimin, Soewarno, Sabaruddin and Susanna 2018 J. Phys. Conf. Ser. 1116 032009.
[29] Guntara Y and Wilujeng I 2018 Jurnal Kependidikan 2, 302-319.
[30] Mondal B C 2013 Int. J. on New Trends in Education and Their Implications 4, 23-34.
[31] Wetherick N E and Davies P 1972 British Journal of Psychology 63, 555–559.
[32] Konya E 2019 The Electronic Journal of Mathematics and Technology 13, 181-189.
[33] Razali, Halim A, Haji A G and Nurfadilla E 2020 J. Phys. Conf. Ser. 1460 012118.
[34] Mizayanti, Halim A, Safitri R and Nurfadilla E 2020 J. Phys. Conf. Ser. 1460 012124.
[35] Halim A, Meerah T S and Halim L 2009 Sains Malaysiana 38, 91-98.
[36] Purwanto N 2001 Prinsip-Prinsip dan Teknik Evaluasi Pengajaran (Bandung: PT Remaja Rosdakarya)
[37] Alfonso V C, Dawn P, Flanagan and Safaa A R 2005 “The Impact of the Cattell-Horn- Carroll Theory on Test Development and Interpretation of Cognitive and Academic Abilities.”. In D. P. Flanagan and P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues, (New York:Guilford Press)
[38] Green D R, Yen W M and Burket G R 1989 Applied Measurement in Education 2, 297-312.
[39] Bichi A A and Talib R 2018 Int. J. of Evaluation and Research in Education 7, 142-151.
[40] Waller J, Ostini R, Marlow L A, McCaffery K and Zimet G 2013 Preventive Medicine 56, 35–40.
[41] Heaton J B 1988 Writing English Language Tests (New York:Longman Inc)
[42] Hake R R 1999 Analyzing Change/Gain Score (Indiana: Indiana University)
[43] Hake R R 1998 Am. J. Phys. 66, 64-76.
[44] Brown H G 1998 Language Assessment; Principle and Classroom Practices (New York: Longman)
[45] Spector P E 1992 Summated Rating Scale Construction An Introduction (Newbury Park: CA Sage)
[46] Sumadi S 1997 Pengembangan tes hasil belajar (Jakarta: PT. Raja Grafindo Persada)

Acknowledgments
We would like to thank all those who have helped carry out the research. Especially lecturers who are in the Department of Physics Education at the Faculty of Teacher Training and Education, Syiah Kuala University.