The Content Validity: Two-Tier Multiple Choices Instrument to Measure Higher-Order Thinking Skills

S R Maulita*, Sukarmin¹, A Marzuki¹

¹Physics Education Department of Graduate Program Sebelas Maret University, Jl. Ir Sutamin 36A Kentingan, Jebres, Surakarta 57126, Indonesia

*septhiriamaulita.93@student.uns.ac.id

Abstract. Two-tier multiple choices instrument is an instrument consisting of two tiers, the first tier is choices of the answer from a question and second tier is choices of the reason for an answer in the first tier. The two-tier instrument can be used to measure students’ higher-order thinking skills. This research aims to analyze content validity two-tier instrument multiple choice to measure students' higher-order thinking skills in a static fluid. The method used in this research is descriptive analyze a method for content validation based on Aikens' Formula. Data of content validation was obtained from 9 experts, consisting of 6 lecturers who had doctorates and 3 physics teachers of the senior high school in Surakarta. Based on the results and discussion, can be concluded that 20 item questions of two-tier multiple choices instrument to measure students' higher-order thinking skills in the static fluid are declared valid based on Aiken Validity.

1. Introduction
In the globalization era, teachers prosecuted to improve students’ higher-order thinking skills. According to Blooms’ taxonomy revised, there are six levels, namely remember, understand, apply, analyze, evaluate, and create. Remember, understand, and application are lower-order thinking skills (LOTS) while analyze, evaluate and create are higher-order thinking skills (HOTS) [1].

Higher Order Thinking Skills as a skill that occurs when someone connects information that stores in memory with new information, then conveys the combined information to reach a goal or answer needed. Empowerment of higher-order thinking skills can be done by giving unusual or uncertain problems [2]. Giving an unusual problem can make question or dilemma, therefore the application is said successful when students succeed to explain, decide, indicate and solving problems in the context of knowledge and experience [3]. Questions of higher-order thinking skills can encourage students to think deeply about the material lesson, so it can be said that question of higher-order thinking skills can provide stimulation to students for developing their higher-order thinking skills [4].

The role of teachers in improving the quality of education can by improving the quality of learning and assessment system [5]. The ability of teachers to master classes that balanced with the ability to conduct judgments is crucial in the context of subsequent learning planning [6, 7]. Assessment in the learning system is an inseparable thing in a learning [8], therefore assessment of student learning outcomes needs to be carried out continuously so that the development of student learning outcomes is monitored [9].

A good assessment system can encourage teachers to improve the quality of learning by creating a good teaching strategy [10]. Therefore, the accuracy of the selection of assessment methods will greatly influence the objectivity and validity of the assessment results so as to produce objective and valid information on the quality of education [11]. Through assessment, the teacher can find out students'
skills, the accuracy of teaching methods, and success of students in achieving the specified competencies [12]. The assessment can be a series of questions that must be answered or orders that must work on students [13]. Questions can contain questions that test students in solving problems. Students must have high-level to answer the question. Students must also to integrate knowledge, understand, and skills and connect them to new situations.

An instrument is a tool used for assessment, that is a tool to measure information on characteristics of an object. The object can be students; skills, attitudes, interests or motivation [14]. In addition to determining the profile of students' skills, the instrument can train students' skills to think at a higher level.

The questions of two-tier multiple choice develop by Treagust from Cuitin University Australia [15]. The form of question consists of two-tier, the first tier is multiple choice questions and the second tier is the reason for choosing answer of the first tier. The question of two-tier multiple choice is similar to traditional multiple choice questions but as the name suggests, two-tier multiple choice contains two tier of interconnected questions [16]. The purpose of the second tier is to encourage students' higher order thinking and reasoning skills. The first tier of question is usually related to the state of knowledge, while the second tier is facilitated to testing students at higher order thinking. Such an instrument will make easier to test students' higher-order thinking than questions multiple choice conventional. The purpose of this question is to help students and teachers to identify student problem so that they can correct errors or difficulties and develop a deeper understanding of the topic.

Two-tier multiple choices instrument can be used to measure higher-order thinking [17]. Superiority question of two-tier multiple choice can use to measure students' cognitive skills at higher-order thinking [18]. The inclusion of reasons at the second tier can use improve higher-order thinking skills and know the skills of students to provide reasons [19].

A good assessment will function appropriately with the purpose of assessment [20] and can improve the quality of assessment results [21]. The exact assessment can encourage students to learn with higher-order thinking [22]. In order assessment to effective, so necessary to develop an instrument that standard and valid [23].

Validity is a way to prove that an instrument developed is a good assessment. Validity comes from the word validity which means the extent of accuracy of an instrument to measuring a function. The instrument can be said high validity if an instrument has a measuring function or provides a measurement result that accordance with purpose the measurement. The instrument that produces data not relevant to the purpose of measurement is said instrument that low validity [21].

According to APA (American Psychological Association) there are three types of validity, namely 1) criterion-related validity, 2) construct validity, and 3) content validity [22]. Content validity is the extent to which items in instrument represent entire contents of components measured and to extend to which items reflect the characteristics of behavior to measured [22]. The determination of content validity can be done by assessing the suitability between the indicator grid and the operational definition of an instrument carried out by at least 5 experts [23]. Thus, an instrument is said valid, according to content validity if instrument represents indicator to measure.

Based on the description above, the content validity of an instrument needs to be done. Therefore, this research will analyze the content validity of the two-tier instrument multiple choice to measure students' higher-order thinking skills. The content validity of the two-tier multiple-choice instrument involved 9 experts. Gusti Ayu also conducted the same research with the aim to obtain an overview of the result of content validity instrument used to test the use of digital modules in discrete mathematics courses. The expert involved by Gusti Ayu in conducting content validity on this research only 2 experts, namely experts in the field of informatics engineering education and experts in the field of mathematics education [24].

2. Research Method

This research is part of the research on the development of two-tier multiple choices instrument to measure higher-order thinking skills. An indicator of higher-order thinking skills used in developing
instrument two-tier multiple choice to measure higher-order thinking skills namely analyze, evaluate and create. Two-tier multiple choices instrument is an instrument that consists of two tiers. The first tier is a question with HOTS indicator and the second tier is the reason for selecting an answer at first tier. Two-tier multiple choices instrument was developed as many as 20 questions to measure students' higher-order thinking skills on the static fluid material.

The development of the instrument multiple choice to measure higher-order thinking skills adapts from research and development (R&D) 4-D model proposed by Thiagarajan [24]. Based on the 4-D model, the stages of research and development of the two-tier multiple choices instrument consist of define, design, develop and disseminate. The stages of two-tier multiple choices instrument can be seen in figure 1.

![Figure 1 Scheme of research and development procedure 4-D model](image)

Figure 1 is a scheme of research and development of instrument multiple choice to measure higher-order thinking skills. In this research, it will be discussed in the stage of develop stage, that is the result of validation two-tier multiple choices instrument to measure higher-order thinking skills that produced draft II.

The research method used is descriptive research method. The object of research is the two-tier multiple choices instrument to measure higher-order thinking skills. The subjects involved in the validation of two-tier multiple choices instrument to measure higher-order thinking skills were 9 experts,
consisting of 6 lecturers who had doctorates and 3 physics teachers of the senior high school in Surakarta.

Validation of instrument is obtained from review expert with 4 categories: irrelevant (IR) if a question does not reflect indicators of the question, indicators of higher-order thinking skills and KD that want to measure. Less relevant (LR) if skills or material that want to measure is not appropriate (need revision). Relevant enough (RE) if an indicator, skills or material that want to measure is appropriate but any error in the words (language) and relevant (R) if an indicator, skills, and material that want to measured are appropriate and construction, language, and material are good [25]. Experts are asked to provide a check mark (√) for every question in the table prepared as shown in figure 2.

In figure 2, the sheet of validation two-tier multiple choices instrument to measure higher-order thinking skills consists of question indicators, HOTS indicators, question and answer, the category of validation and suggestion column. Through a sheet of validation, the expert can see the relationship between question indicators, question and HOTS indicator. The expert provides a check mark (√) in the appropriate category. Whereas in the suggestion column, the expert can give suggestions regarding question being reviewed.

Validation consists of 4 categories, namely irrelevant (IR) worth 1, less relevant (LR) worth 2, relevant enough (RE) worth 3 and relevant (R) worth 4. The result of validation each item from 9 experts were analyzed using Aiken Formula [26], then obtained the value of content validity of two-tier multiple choices instrument to measure students’ higher-order thinking skills. Aiken formula that used to calculate the content validity coefficient based on the result of validation an item as many n people, the extent to which the item represents the measured indicators is as follows [20]:

$$ V = \frac{2S}{n(C-1)} $$

$$ s = r - L_o $$
where:
V = index of validity Aiken
\( L_o \) = lowest value/category (in this category is irrelevant (IR) worth 1)
C = highest value/category (in this category is relevant (R) worth 4)
r = value/category from expert
n = number of expert

**Table 1** Example for calculation of contents validity

| Expert | value (r) | \( s = r - L_o \) |
|--------|----------|-------------------|
| 1      | 3        | 3 - 1 = 2         |
| 2      | 4        | 4 - 1 = 3         |
| 3      | 3        | 3 - 1 = 2         |
| 4      | 4        | 4 - 1 = 3         |
| 5      | 4        | 4 - 1 = 3         |
| 6      | 3        | 3 - 1 = 2         |
| 7      | 4        | 4 - 1 = 3         |
| 8      | 4        | 4 - 1 = 3         |
| 9      | 4        | 4 - 1 = 3         |
| \( \Sigma s \) |          | 24                |
| V      |          | 0.889             |

Table 1 is an example for calculation of item number 1 using Aiken Formula. Each question is analyzed using the same method. Value of \( V_{obtain} \) fanged from 0-1. Criteria used to declare a question said to valid in content validity with 9 experts based on Aiken table is 0.74 [26].

3. Result and Discussion
The results of the development two-tier multiple choices instrument to measure higher-order thinking skills consist of five answer choices in the first tier and five reasoning choices in the second tier. Figure 3 is an example of a question developed.

![Figure 3: Example of a question developed](image)
The result of validation from 9 experts was calculated using Aiken Formula with the help of the Microsoft Excel program. Each item of questions has a value of validation from 9 experts. The result of calculated validation is shown in table 2.

| Number Of Question | V obtained | V table | Conclusion |
|--------------------|------------|---------|------------|
| 1                  | 0.93       | 0.74    | Valid      |
| 2                  | 0.96       | 0.74    | Valid      |
| 3                  | 0.85       | 0.74    | Valid      |
| 4                  | 0.89       | 0.74    | Valid      |
| 5                  | 0.93       | 0.74    | Valid      |
| 6                  | 0.81       | 0.74    | Valid      |
| 7                  | 1.00       | 0.74    | Valid      |
| 8                  | 0.96       | 0.74    | Valid      |
| 9                  | 0.93       | 0.74    | Valid      |
| 10                 | 1.00       | 0.74    | Valid      |
| 11                 | 0.96       | 0.74    | Valid      |
| 12                 | 1.00       | 0.74    | Valid      |
| 13                 | 0.96       | 0.74    | Valid      |
| 14                 | 1.00       | 0.74    | Valid      |
| 15                 | 1.00       | 0.74    | Valid      |
| 16                 | 0.96       | 0.74    | Valid      |
| 17                 | 0.96       | 0.74    | Valid      |
| 18                 | 0.96       | 0.74    | Valid      |
| 19                 | 0.96       | 0.74    | Valid      |
| 20                 | 1.00       | 0.74    | Valid      |

Values of Vtable in table 2 are obtained from the Aiken Table. The instrument validated by 9 raters with 4 validation categories will be valid if the results of the validation calculation are least 0.74 [26]. Based on the calculation of content validity with Aiken formula, the results are $V_{obtained} > V_{table}$, so that 20 item questions that developed can be declared valid. Thus 20 item questions are suitable to measure indicators of higher-order thinking skills that have been determined based on content validity.

Validation by experts can be proof of validity. According to Aiken, there is nothing wrong in using expert judgment if the assessment is made carefully and independently[26]. Retnawati also stated that content validity can be determined using expert agreement in the field, which is stated valid if the expert believes that the instrument can measure what should be measured [27]. The expert’s task is to look at the suitability of the indicator with the purpose of developing instrument, the suitability of indicator with scope of the material or the suitability of theory, seeing the suitability of the instrument with item indicators, seeing the truth of the concept questions, key truths, language and culture [27]. Therefore the selection of experts is very important because it will determine the quality of the instrument being developed.

4. Conclusion

Based on the results of research and discussion, can be concluded that 20 item questions of two-tier multiple choices instrument to measure students’ higher-order thinking skills in the static fluid are declared valid based on Aiken Validity. The two-tier multiple choices instrument represents indicators to measure higher-order thinking skills.
Acknowledgments
The authors thank her parents for their continuous prayers and special thanks to Mr. Sukarmin S.Pd., M.Si. Ph.D and Mr. Ahmad Marzuki, M.Si. Ph.D. because this paper would have been completed without the encouragement and counsel of them.

References
[1] H. Saputra 2016 Pengembangan Mutu Pendidikan Menuju Era Global (Jakarta: CV Smile Indonesia Institute)
[2] A. Lewis & D. Smith 2001 Defining higher order thinking Theory Pract. 32 3 131–137
[3] F. King, L. Goodson & F. Rohani 2010 Higher Order Thinking Skills Definition Teaching Strategies Assessment A Publ. Educ. Serv. Program, now known as Cent. Adv. Learn. Assess. www.cala.fsu.edu, 18 458
[4] J. E. Barnett & A. L. Francis 2012 Using higher order thinking questions to foster critical thinking: A classroom study Educ. Psychol. 32 2 201–211
[5] M. Hasyim 2014 Penerapan Fungsi Guru Dalam Proses Pembelajaran Auladuna. 1 2 269
[6] Abdurrarahman 2017 Efektivitas dan Kendala Pembelajaran Sains Berbasis Inkuiri terhadap Capaian Dimensi Kognitif Siswa : Meta Analisis. Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah. 2 1 1–9 https://doi.org/https://doi.org/10.24042/tadris.v2i1.1206
[7] S. Suranto, M. Muhaydi & D. Mardapi 2014 Pengembangan instrumen evaluasi uji kompetensi keahlian (UKK) administrasi perkantoran di SMK J. Penelit. dan Eval. Pendidik. 18 1 98–114
[8] K. Büyükkarac 2014 Assessment Beliefs and Practices of Language Teachers in Primary Education Int. J. Instr. 7 1 107-121
[9] Supahar & Z. K. Prasetyo 2015 Pengembangan Instrumen Penilaian Kinerja Kemampuan Inkuiri Peserta Didik Pada Mata Pelajaran Fisika SMA J. Penelit. dan Eval. Pendidik., 19 1 97
[10] D. Mardapi 2012 Pengukuran Penilaian dan Evaluasi Pendidikan. Yogyakarta: Nuha Litera
[11] H. Setiadi 2016 Pelaksanaan penilaian pada kurikulum 2016 J. Penelit. dan Eval. Pendidik. 20 2 166–178
[12] A. S. Shidiq, M. Masykuri & E. S. Van Hayes 2014 Pengembangan Instrumen Penilaian Two-Tier Multiple Choice Untuk Mengukur Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking Skills) Pada Materi Kelarutan Dan Hasil Kali Kelarutan Untuk Siswa Sma/Ma Kelas XI J. Pendidik. Kim. 3 4 83–92
[13] A. Sudijono 2009 Pengantar Statistik Pendidikan (Jakarta: Rajawali)
[14] E. P. Widoyoko 2012 Teknik Penyusunan Instrumen Penelitian (Yogyakarta: Pustaka Pelajar).
[15] D. Treagust 2012 Diagnostic assessment in science as a means to improving teaching, learning and retention Aust. Conf. Sci. 1998 1–9
[16] A. Cullinane & M. Liston 2011 Two-tier Multiple Choice Question: An Alternative Method of Formative Assessment for First Year Undergraduate Biology Students (Limerick: National Center for Excellence In Mathematics and Education Science Teaching and Learning (NCE-MSTL))
[17] Basuki, Ismet & Hariyanto 2015 Asesmen Pembelajaran (Bandung: Remaja Rosdakarya)
[18] E. Rofiah, N. S. Aminah & E. Y. Ekawati 2013 Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika pada Siswa SMP J. Pendidik. Fis. Univ. Sebel. Maret. 1 2 17–22
[19] E. Istiyono, D. Mardapi & Suparno 2014 Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PysTHOTS) Peserta Didik SMA J. Penelit. Dan Eval. Pendidik. 18 1 1–12
[20] V. Serevina, C. Dewi & D. Muliyati 2016 Rancangan Tes dan Evaluasi Fisika yang Informatif dan Komunikatif pada Materi Kinematika Gerak Lurus JPPPF-Jurnal Penelit. Pengemb. Pendidik. Fis. 2 1
[21] S. Azwar 1995 Reliabilitas Dan Validitas Aitem Bul. Psikol. 3 1 19–26
[22] Budiyono 2015 Pengantar Penilaian Hasil Belajar (Surakarta: UNS Press)
[23] M. J. Wantah 2010 Validitas dan Reliabilitas Instrumen Asesmen Anak TK berkebutuhan Khusus J. Penelit. dan Eval. Pendidik
[24] G. A. D. Sugiharni 2017 Validitas Isi Instrumen Pengujian Modul Digital Matematika Diskrit Berbasis Open Source di STIKOM Bali Konferensi Nasional Sistem & Informatika. 678–684
[25] S. Thiagarajan, D. S. Semmel & M. I. Semmel 1974 Instructional Development for Training Teachers of Exceptional Children: A Sourcebook, no. Mc
[26] A. S. Shidiq 2016 Pengembangan Instrumen Penilaian Teslet Untuk Mengukur Keterampilan Proses Sains Pada Materi Hidrolisis Garam Untuk Siswa Kelas XI SMA/M Universitas Sebelas Maret
[27] L. R. Aiken 1980 Content validity and reliability of single items or questionnaires,” Educ. Psychol. Meas. 40 4 955–959
[28] H. Retnawati 2016 Analisis Kuantitatif Instrumen Penilaian (Panduan Peneliti, Mahasiswa, dan Psikometrian) (Yogyakarta: Parama Publishing)