**Journal of Education and Learning (EduLearn)**
Vol. 19, No. 1, February 2025, pp. 551~560
ISSN: 2089-9823 DOI: 10.11591/edulearn.v19i1.21207 

# Proving content validity of android-based higher order thinking skill assessment for science and mathematics preservice teacher

**Endang Susantini[1], Yurizka Melia Sari[2], Prima Vidya Asteria[3], Muhammad Ilyas Marzuqi[4]**
[1]Department of Biology Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, Surabaya, Indonesia
[2]Department of Mathematics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, Surabaya, Indonesia
[3]Department of Indonesian Language Education, Faculty of Languages and Arts, Universitas Negeri Surabaya, Surabaya, Indonesia
[4]Department of Social Sciences Education, Faculty of Social Sciences and Law, Universitas Negeri Surabaya, Surabaya, Indonesia

**Article Info**

***Article history:***

Received Aug 19, 2023
Revised Mar 28, 2024
Accepted May 18, 2024

***Keywords:***

Aiken validity
Assessment
Content validity
Education
Higher order thinking skill
Mathematics
Science

**ABSTRACT**

Assessing preservice' higher order thinking skills (HOTS) in science and mathematics is essential. Teachers' HOTS ability is closely related to their ability to create HOTS-type science and mathematics problems. Among various types of HOTS, one is Bloomian HOTS. To facilitate the preservice teacher to create problems in those subjects, an Android app called EduAssess was developed as a Bloomian HOTS test for junior high school preservice teachers. This study aims to validate the problems in the EduAssess app through content validity. Content validity was analyzed using Aiken's V formula and expanded Gregory formula. EduAssess comprised three test sets for science and mathematics, each comprising 9 items. The instrument validated by three experts in each subject. The study results demonstrate that EduAssess, for both mathematics and science, has achieved content validity. Expert judgments confirmed the validity of EduAssess items, with Aiken's V index ranging from 0.67 to 1.00, meanwhile expanded Gregory index ranging from 0.78-1.00. The results showed that EduAssess includes analysis, evaluation, and creation. The findings highlight that the application instrument in facilitating pre-service teachers by measuring their ability to analyze, evaluate and create HOTS problems in science and mathematics subjects was proven valid and ready for data collection.



***Corresponding Author:***

Yurizka Melia Sari
Department of Mathematics Education, Faculty of Mathematics and Natural Sciences
Universitas Negeri Surabaya
Ketintang Street, Gayungan, Surabaya, Indonesia
Email: yurizkasari@unesa.ac.id

## 1. INTRODUCTION

An endeavour to improve the quality of education is through curriculum development that can serve as a reference in the learning process. Based on the Decree of the Minister of Education, Culture, Research, and Technology Number 56/M/2022 [1] regarding guidelines for implementing the curriculum in learning recovery, it is stated that it is currently necessary to restore learning that has experienced learning loss due to the pandemic [2]. Therefore, educational institutions and teachers must develop a curriculum that meets the conditions of educational institutions, the local area's potential, and the student's needs. The intended curriculum is the Merdeka Curriculum, with the concept of "Merdeka Belajar" for students. The Merdeka Curriculum is an initiative of the Indonesian government to improve the quality of education in Indonesia through learner-centered learning and paying attention to cultural diversity and local potential in each region

[1]‒[4]. In the Merdeka Curriculum, teachers must develop students' competencies through creative and innovative learning.

One of the creative and innovative forms of learning is by developing problems based on higher order thinking skills (HOTS) [5]. HOTS is a high-level thinking ability to analyze, synthesize, evaluate, and apply information or knowledge in new or complex situations [6]. High-level thinking is the ability to think beyond remembering, rewriting, and referring without processing information, but the ability to examine problems critically and creatively. Indicators of high-level thinking skills, according to Krathwohl [7], are i) the ability to analyze, which consists of the ability to distinguish the cause and effect of a problem and identify a statement, ii) the ability to evaluate, which includes making hypotheses, and conducting tests, and iii) the ability to create in the form of activities to design a simple experiment to solve a problem [8].

HOTS problems that can improve students' critical and creative thinking skills still need to be improved in junior high school students. It is because the ability of junior high school teachers to create HOTS problems still needs to be improved [9]‒[13]. It can be seen from the results of the assessment of HOTS problems made by teachers that only a small number of problems are categorized as HOTS problems. These results are also supported by other research [11] that stated HOTS problems made by junior high school teachers in essays and multiple choice still need to be higher. It is evident from the total number of problems still in memorizing category (C1). Therefore, junior high school teachers must be able to make HOTS problems in the Merdeka Curriculum as part of efforts to improve the quality of learning and create a generation capable of critical and creative thinking [14].

An important strategy to improve teachers' skills in developing HOTS problems is through regular and continuous training and mentoring [15]. This approach not only creates opportunities for teachers to improve their understanding of HOTS concepts continuously but also provides a platform to hone their practical skills in designing problems that meet the demands of the independent curriculum [16]. Regularly organized training has significant benefits in equipping teachers with a deeper understanding of the basic principles of HOTS. In a dynamic learning environment, teachers must have a strong insight into what types of problems can stimulate higher-order thinking in students [17]. Through training, teachers can gain concrete guidance on designing problems that lead to analysis, evaluation, and creation skills and integrating these key aspects into an independent curriculum.

One of the forms of training conducted is technology-integrated training so that it can accommodate many participants. Some research related to the development of digital assessment integrated with Bloom's taxonomy provides positive results [18], [19]. The study shows that Bloom's Cognitive Domain is connected to Bloom's taxonomy system digitally, which is integrated with information communication technology devices to create online assessments that can meet the diverse ability needs of students during the pandemic through online planning, teaching, and learning methods. In addition, Susantini et. al. [20] research related to the development of an assessment link application to improve the ability of prospective Biology teacher students to create HOTS problems also gave positive results and had a good impact. It shows that the urgency of developing digital assessments that help teachers make problems according to Bloom's taxonomy has yet to be done, especially at the junior high school level.

An innovative solution is needed to improve teachers' ability to create quality HOTS problems. To facilitate teachers, Android-based applications emerge as an attractive alternative. Such applications can assist teachers in developing HOTS problems following the characteristics of the independent curriculum. The interactivity and sophistication of Android technology can facilitate the process of developing problems so that teachers can focus more on designing problems that encourage students to think critically and creatively. This study investigated content validity instrument on Android based application to measure teacher ability to create HOTS problems. While earlier studies have explored the impact of technology on student's HOTS. They have not explicitly addressed its influence on content validity ensures that the problems in the application are truly capable of measuring HOTS. Therefore, the research conducted is to prove that the content validity of Android-based HOTS assessment instruments is important in developing prospective teachers' abilities in developing HOTS problems.

## 2. METHOD

This research was carried out within the framework of research and development (RnD), aiming to create an Android-based electronic assessment application to assist aspiring junior high school science and mathematics teachers in formulating HOTS challenges. According to this objective, our investigation centred on establishing the validity of this assessment tool. The principal method for analysing data in this study was content validity. To prove the validity of the assessment instrument, two approaches to content validity were employed the Aiken method and the expanded Gregory method [21].

Content validity is established through the consensus of experts. This consensus, often called the measured domain, establishes the categorization of content validity (content-related). This determination is based on measuring instruments being deemed valid if experts concur that these instruments accurately gauge the proficiencies outlined within the domain or the psychological constructs under scrutiny. In comprehending this concordance, a validity indicator can be employed, encompassing the metric Aiken suggested Aiken's formula for the item validity index [22] is articulated as follows:

$$v = \frac{\sum s}{n(c-1)} \quad (1)$$

where V represents the index of item validity; s denotes the scores allocated by each rater, subtracted by the lowest score present within the designated category (s = r - lo), wherein r signifies the rater's category selection score, and lo symbolizes the least scores within the scoring category); n pertains to the count of raters; and c signifies the count of categories that raters can opt for. From the calculation of the V index, an item can be categorized based on its index. If the index is less than or equal to 0.4, it is said to have less validity, 0.4-0.8 is said to have moderate validity, and if it is more significant than 0.8, it is said to be very valid.

The indicator of expert concurrence regarding content validity entails contrasting the quantities of items validated by three experts, which exhibit substantial pertinence to the wider category of items [23], [24]. The outcomes of the relevance analysis, presented in the form of contingency tables, can be found in Table 1, whereas the formula for the validity coefficient is provided as Formula 2.

$$\text{Content Validity Index from Expanded Gregorian} = \frac{H}{(A+B+C+D+E+F+G+H)} \quad (2)$$

The expanded Gregorian validity index also ranges from 0-1. The closer the validity index value is to 1, the stronger the agreement from experts regarding the content validity of the produced instrument.

Table 1. Table of contingencies to use the Gregory formula and three expert conclusions to determine the validity coefficient

| The experts | 1st Combination | 2nd Combination | 3rd Combination | 4th Combination | 5th Combination | 6th Combination | 7th Combination | 8th Combination |
|---|---|---|---|---|---|---|---|---|
| Expert 1 | Weak | Weak | Weak | Weak | Strong | Strong | Strong | Strong |
| Expert 2 | Weak | Weak | Strong | Strong | Weak | Weak | Strong | Strong |
| Expert 3 | Weak | Strong | Weak | Strong | Weak | Strong | Weak | Strong |
| Total | A | B | C | D | E | F | G | H |

## 3. RESULTS AND DISCUSSION

### 3.1. The mathematics higher order thinking skill instruments

Developing HOTS problems in mathematics for junior high school involves three levels of Bloom's taxonomy: analysis, evaluation, and creation. The main objective is to promote critical and creative thinking and evaluative skills in students 7 to 9. In total, 27 HOTS items were designed and developed. The following are examples of some HOTS item designs in junior high school mathematics subject areas as shown in Table 2.

Developing mathematics HOTS items in the next stage involved careful evaluation and revision. Some items underwent revision after the expert judgment stage, where the opinions and inputs from experts were carefully evaluated and applied to the developed problems. This step ensures that the assessment objectives set in designing HOTS problems are achieved and improves the construction of problems to fit the analysis, evaluation, and creation levels. The results of the example revisions that have been made can be seen in Table 3.

In the analysis aspect, the main objective is to get students to break down more complex math problems into smaller components. It helps students develop their analytical skills. These problems require students to identify patterns, connect concepts, or distinguish relevant parts of a given mathematics problem [25]. The format used is multiple choice, which allows students to choose the answer that best suits their analysis [26]. Meanwhile, in the evaluation aspect, the main focus is getting students to take a critical approach to the mathematical information or problem. It involves students' ability to assess the strengths and weaknesses of an approach or solution [27]. The evaluation problems are also designed in a multiple-choice format, which asks students to consider a given argument and choose the answer closest to their critical evaluation [16]. Lastly, in the creation aspect, students are encouraged to develop new solutions or concepts to overcome the given mathematical challenge. It involves students' ability to think creatively and apply learned mathematical concepts. The creation problems are presented in essay form, allowing students to explain their thinking in depth and detail the steps or processes they took in creating the new solution or concept [8]. By combining these three levels of Bloom's taxonomy for each grade, the development of HOTS problems in junior high school mathematics has created various problems that encourage students to develop HOTS. From analysis to creation, multiple choice to the essay format, these problems challenge students to understand mathematics more deeply and apply their knowledge in different contexts [28].

Table 2. The examples of mathematics HOTS problems developed in grade 8 before revision

| Level | The goal of the assessment | The problems | The format |
|---|---|---|---|
| Analysis | Analyze contextual problems related to the system of linear equations of two variables. | Aldi bought two shirts and one jacket for a total of $Rp$ 138,000.00. Arriving home, he regretted his purchase and exchanged one shirt for another jacket. Aldi can do this, but he must pay $Rp$ 12,000.00 more because the jacket costs more than the T-shirt. If, from the beginning, Aldi had bought one shirt and two jackets, how much would Aldi have to pay?<br>A. IDR 148,000.00<br>B. IDR 138,000.00<br>C. IDR 128,000.00<br>D. IDR 118,000.00 | Multiple choice |
| Evaluation | Evaluate the concept of a system of linear equations of two variables used to solve a given contextual problem. | A restaurant serves several food packages.<br>• Menu A (chicken + iced tea) = Rp 10,000<br>• Menu B (2 chicken + iced tea) = Rp 18,000<br>• Menu C (5 chicken + 2 iced tea) = Rp 45,000<br>Free rice is not included in the package. The price of drinks outside the package, namely iced tea $Rp$ 3,000.00.<br>Auni will treat her friends to eat at the place; Auni has 100,000 to treat 8 of her friends (including herself); which package do you think Auni should choose so that Auni still has plenty of change and gets the most favorable price?<br>A. 1 Menu C + 2 Packages Menu A<br>B. 1 Menu C + 1 Menu B<br>C. 4 Packages Menu B<br>D. 2 Packages Menu C | Multiple choice |
| Creation | Make decisions from contextual problems using the system of linear equations of two variables. | At a clothing store, several discounts are being applied, namely:<br>• Every purchase of 2 trousers free 1 long sleeve T-shirt<br>• Every purchase of 3 pairs of trousers is 30% off.<br>• For the purchase of 1 long sleeve shirt and 1 trousers 50% discount<br>• Every purchase above Rp 250,000 is discounted by 20%<br>If the regular price of 1 pair of trousers is IDR 70,000, 1 short sleeve shirt is IDR 50,000, and 1 long sleeve shirt is IDR 60,000. Auni wants to buy 2 sets of clothes, namely 2 trousers, 1 short sleeve shirt, and 1 long sleeve shirt. What should Auni do to get the cheapest price? | Essay |

Table 3. The example of revision of mathematics HOTS problems

| Level | Before revision | Some expert judgement comment | After revision |
|---|---|---|---|
| Analyze | Aldi bought two shirts and one jacket for a total of IDR 138,000.00. Arriving home, he regretted his purchase and exchanged one shirt for another jacket. Aldi can do this, but he must pay IDR12,000.00 more because the jacket costs more than the T-shirt. If, from the beginning, Aldi had bought one shirt and two jackets, how much would Aldi have to pay?<br>A. IDR148,000.00<br>B. IDR138,000.00<br>C. IDR128,000.00<br>D. IDR118,000.00 | The created problem is not suitable for the analysis level because, without direct calculation, it can be found by adding directly. Conditions must be added to make this problem more complex and can fit the analysis stage. | Aldi bought two shirts and one jacket at shop A for IDR138,000. Arriving home, he regretted his purchase and exchanged one shirt for another jacket. Aldi can do this, but he must pay IDR12,000 more because the jacket costs more than the shirt. On the same day, Bima bought one shirt and jacket at shop A. How much should Aldi pay?<br>A. IDR90,000<br>B. IDR92,000<br>C. IDR94,000<br>D. IDR96,000 |

### 3.2. The natural science higher order thinking skill instruments

In the research conducted on developing HOTS problems in the subject of natural sciences for junior high school, 27 problems were successfully developed. These problems cover three aspects of Bloom's taxonomy, namely analysis, evaluation, and creation, and are intended for grades 7, 8, and 9. Problems related to the analysis and evaluation aspects are designed as multiple choice, while problems related to the creation aspect are presented as essays. While developing HOTS problems for Natural Science subjects at the junior high school level, it was found that some items needed further improvement. Based on the assessment of one of the experts, some of the sentences in the problems were considered ambiguous, and although they had been designed as contextual problems, the realization of the context in the problems needed to be considered more. It indicated that adjustments needed to be made so that the problems could more effectively measure the desired HOTS. Details of the improvements that need to be made have been carefully collated and can be found in Table 4.

Table 4. The example of revision of Natural Science HOTS problems

| Level | Before revision | Some expert judgement comment | After revision |
|---|---|---|---|
| Analyze | Take a look at the following picture.<br>Iron  Plastics<br>Two spoons made of iron and plastic are dipped in a glass filled with hot water. After a while, the temperature of the two spoons is measured using a thermometer. The statement below that is correct regarding the temperature of the two spoons is<br>a. The temperature of the iron spoon is lower than the plastic spoon because the iron spoon receives more heat.<br>b. The temperature of the iron spoon is lower than that of the plastic spoon because the iron spoon receives less heat.<br>c. The temperature of the iron spoon is higher than that of the plastic spoon because the plastic spoon receives more heat.<br>d. The temperature of the metal spoon is the same as that of the plastic spoon because both spoons received the same amount of heat. | This problem item is potentially ambiguous. If an iron spoon is dipped in hot water for the same amount of time, it will be hotter than a plastic spoon for the same amount of time. Another less contextual thing is how to measure a spoon's temperature with a thermometer.<br><br>In the explanation, there are two concepts: thermal interaction and conductor/isolator of heat. It is better to focus on what this problem is related to, whether it is a conductor/insulator of heat or an insulator of heat. | Take a look at the following picture.<br>Iron  Plastics<br>Two spoons made of iron and plastic are dipped in a glass containing hot water. After a while, there is thermal equilibrium. The two spoons are measured using a thermometer. The following statement is correct regarding the temperature of the two spoons...<br>a. The temperature of the iron spoon is lower than the plastic spoon because the iron spoon receives more heat.<br>b. The temperature of the metal spoon is lower than that of the plastic spoon because the metal spoon receives less heat.<br>c. The temperature of the iron spoon is higher than that of the plastic spoon because the plastic spoon receives more heat.<br>d. The temperature of the metal spoon is the same as that of the plastic spoon because both spoons received the same amount of heat. |

In developing the design of science problems, the judgment feedback emphasized the importance of using precise and clear sentences. The main objective was to ensure that the science concepts tested were clear for students. A common misconception is that HOTS problems must be deliberately ambiguous to increase their difficulty level. However, HOTS problems should test students' HOTS without confusing them with unclear problem formulations [11], [29]. Therefore, it is important for such problems to be presented with clarity and precision so that students can focus on understanding concepts and applying critical thinking in answering problems rather than on trying to understand ambiguous problems.

### 3.3. Proving mathematics and science content validity using Aiken formula

In the content validation process for the mathematics assessment instrument using Aiken's formula, validity results ranging from 0.67 to 0.89 were obtained as shown in Table 5. It indicates that the instrument developed has a very high level of validity, reflecting the accuracy and suitability of the items to the domain of mathematics concepts to be measured. The validation process involved the assessment of three expert judgments: a lecturer specializing in the evaluation of mathematics education, a lecturer in mathematics, and a lecturer in mathematics learning. The involvement of these three experts ensured that the validation was comprehensive, considering aspects of evaluation, mathematical content and learning approaches.

Compared to the validity results for the mathematics instruments, the Science instruments showed a similar range of validity, between 0.78 and 1.00, when measured by Aiken's formula as shown in Table 6. Although the numbers are similar, it is important to remember that the two fields of study have different nuances and complexities [21]. The validation process of the science instruments involved three expert judgments with specific expertise in their fields. The assessment was conducted by a lecturer specializing in science education evaluation, a lecturer in Physics Education, and a lecturer in Biology learning. The three of them ensured that each item in the instrument reflected the science concepts appropriately and relevantly, following the applicable curriculum and educational standards.

Table 5. The Results of content validity of mathematics HOTS instrument using Aiken formula

| Level | Items | Expert 1 | Expert 2 | Expert 3 | Validity Index | The average validity |
|---|---|---|---|---|---|---|
| Analysis | 1 | 4 | 3 | 4 | 0.89 | |
| | 2 | 4 | 3 | 3 | 0.78 | |
| | 3 | 3 | 3 | 3 | 0.67 | |
| | 4 | 4 | 2 | 4 | 0.78 | |
| | 5 | 4 | 3 | 4 | 0.89 | 0.80 |
| | 6 | 4 | 3 | 4 | 0.89 | |
| | 7 | 4 | 3 | 4 | 0.89 | |
| | 8 | 4 | 4 | 2 | 0.78 | |
| | 9 | 3 | 3 | 3 | 0.67 | |
| Evaluation | 10 | 4 | 4 | 3 | 0.89 | |
| | 11 | 4 | 4 | 3 | 0.89 | |
| | 12 | 3 | 3 | 4 | 0.78 | |
| | 13 | 4 | 3 | 4 | 0.89 | |
| | 14 | 4 | 3 | 4 | 0.89 | 0.86 |
| | 15 | 3 | 3 | 4 | 0.78 | |
| | 16 | 4 | 3 | 4 | 0.89 | |
| | 17 | 4 | 3 | 4 | 0.89 | |
| | 18 | 4 | 3 | 4 | 0.89 | |
| Create | 19 | 4 | 3 | 4 | 0.89 | |
| | 20 | 3 | 4 | 4 | 0.89 | |
| | 21 | 4 | 4 | 3 | 0.89 | |
| | 22 | 4 | 4 | 3 | 0.89 | |
| | 23 | 4 | 3 | 4 | 0.89 | 0.86 |
| | 24 | 4 | 3 | 4 | 0.89 | |
| | 25 | 4 | 3 | 3 | 0.78 | |
| | 26 | 4 | 3 | 3 | 0.78 | |
| | 27 | 4 | 3 | 4 | 0.89 | |

Table 6. The results of content validity of science HOTS instrument using Aiken formula

| Level | Items | Expert 1 | Expert 2 | Expert 3 | Validity index | The average validity |
|---|---|---|---|---|---|---|
| Analysis | 1 | 4 | 4 | 4 | 1.00 | |
| | 2 | 4 | 4 | 3 | 0.89 | |
| | 3 | 3 | 4 | 3 | 0.78 | |
| | 4 | 3 | 4 | 4 | 0.89 | |
| | 5 | 4 | 3 | 4 | 0.89 | 0.91 |
| | 6 | 4 | 4 | 4 | 1.00 | |
| | 7 | 4 | 4 | 4 | 1.00 | |
| | 8 | 4 | 4 | 4 | 1.00 | |
| | 9 | 4 | 4 | 2 | 0.78 | |
| Evaluation | 10 | 4 | 4 | 3 | 0.89 | |
| | 11 | 4 | 4 | 3 | 0.89 | |
| | 12 | 4 | 4 | 2 | 0.78 | |
| | 13 | 4 | 3 | 4 | 0.89 | |
| | 14 | 4 | 4 | 4 | 1.00 | 0.88 |
| | 15 | 3 | 3 | 4 | 0.78 | |
| | 16 | 4 | 2 | 4 | 0.78 | |
| | 17 | 4 | 4 | 4 | 1.00 | |
| | 18 | 4 | 3 | 4 | 0.89 | |
| Create | 19 | 4 | 3 | 4 | 0.89 | |
| | 20 | 4 | 4 | 4 | 1.00 | |
| | 21 | 4 | 4 | 3 | 0.89 | 0.90 |
| | 22 | 4 | 4 | 2 | 0.78 | |

In this study, the content validity of the HOTS items for mathematics and science was measured using Aiken's formula. We found that that all items have a validity value of more than 0.6. It indicates that the HOTS problems developed for mathematics and science have correlate with the standard of good content validity [30], [31]. Thus, the problems can be ensured to appropriately represent important concepts in mathematics and science and follow the measurement objectives.

In developing HOTS problems for mathematics, there were significant differences in content validity between the different aspects of Bloom's taxonomy. Interestingly, the content validity for the analysis aspect is lower than evaluate and creation. Meanwhile, both aspects have the same content validity for evaluation and creation. In contrast, in the HOTS problems for science, the content validity for the aspect of evaluation is the lowest, while create aspect reaches the highest value. This difference may be due to the difficulty in formulating analysis problems in mathematics that represent concepts, while in science, evaluation problems may only partially represent the expected complexity and depth of the material [7]. Our study suggests that creating requires deep understanding and synthesis of information in learning, so when problems are well formulated, their validity tends to be higher. It points to the importance of continuous revision and evaluation to improve problem quality and ensure optimal content validity [32], [33].

### 3.4. Proving mathematics and science content validity using expanded Gregorian formula

In the validation process of the mathematics assessment instrument, this study also used the Expanded Gregorian Index method to measure the content validity of the problems that had been developed. The results obtained showed a validity range between 0.78 to 1.00 as shown in Table 7. In the context of developing HOTS problems for natural Sciences, it was found that the content validity results were similar to those of mathematics problems. Using the Expanded Gregorian Index as the measurement method, the range of science content validity was between 0.78 and 0.89 as shown in Table 8.

Table 7. The results of content validity of mathematics HOTS instrument using expanded Gregorian formula

| The experts | 1st Combination | 2nd Combination | 3rd Combination | 4th Combination | 5th Combination | 6th Combination | 7th Combination | 8th Combination | Validity index |
|---|---|---|---|---|---|---|---|---|---|
| Expert 1 | Weak | Weak | Weak | Weak | Strong | Strong | Strong | Strong | |
| Expert 2 | Weak | Weak | Strong | Strong | Weak | Weak | Strong | Strong | |
| Expert 3 | Weak | Strong | Weak | Strong | Weak | Strong | Weak | Strong | |
| Analyze | - | - | - | - | - | 1 | 1 | 7 | 0.78 |
| Evaluate | - | - | - | - | - | - | - | 9 | 1.00 |
| Create | - | - | - | - | - | - | - | 9 | 1.00 |
| Total | - | - | - | - | - | 1 | 1 | 25 | 0.93 |

Table 8. The results of content validity of science HOTS instrument using expanded Gregorian formula

| The experts | 1st Combination | 2nd Combination | 3rd Combination | 4th Combination | 5th Combination | 6th Combination | 7th Combination | 8th Combination | Validity index |
|---|---|---|---|---|---|---|---|---|---|
| Expert 1 | Weak | Weak | Weak | Weak | Strong | Strong | Strong | Strong | |
| Expert 2 | Weak | Weak | Strong | Strong | Weak | Weak | Strong | Strong | |
| Expert 3 | Weak | Strong | Weak | Strong | Weak | Strong | Weak | Strong | |
| Analyze | - | - | - | - | - | - | 1 | 8 | 0.89 |
| Evaluate | - | - | - | - | - | 1 | 1 | 7 | 0.78 |
| Create | - | - | - | - | - | - | 1 | 8 | 0.89 |
| Total | - | - | - | - | - | 1 | 3 | 23 | 0.85 |

Based on the analysis using the expanded Gregorian Formula, it was found that in the mathematics HOTS questions, the content validity for the analysis aspect was lower compared to evaluation and creation. Interestingly, both aspects show the same content validity for evaluation and creation. In contrast, in the science HOTS questions, content validity for the evaluation aspect was the lowest, while analysis and creation recorded the highest content validity values. Overall, these results show no significant difference compared to the calculations made using the Aiken validity index. The accuracy of both methods in providing similar results may be due to their strong conceptual foundations in measuring content validity [21], [30], [31]. Although each has different approaches and nuances, both focus on assessing the congruence between question items and the concept domain to be measured. This study explored a comprehensive difference in validity results between aspects may be due to variations in the complexity and difficulty in formulating questions for each aspect, as well as how the questions represent key concepts from maths and science. However, further and in-depth studies may be needed to confirm its validity result especially regarding the content and the methods. Future research could explore other approaches to measure content and construct validity in ways that are feasible for producing technology-integrated HOTS instruments.

## 4. CONCLUSION

Based on the results and discussion, the final product of this research is a valid HOTS assessment for prospective mathematics and science teachers integrated with an Android application. This instrument is part of the EduAssess application development research used to help prospective mathematics and science teachers to distinguish HOTS and LOTS problems. The assessment instrument for each field of study consists of 27 HOTS problems declared to have appropriate content validity. The developed mathematics instrument has an Aiken validity coefficient of 0.67-0.89, while the science instrument has a higher Aiken validity coefficient of 0.78-1.00. The results of the validity of the mathematics and science content obtained with Aiken are also not much different from the expanded Gregorian formula index. Therefore, the developed instruments can be used as an accurate assessment tool for HOTS in mathematics and science. Our findings provide conclusive evidence that the proposed method of proving content validity is related to expert judgment rather than the result of a limited user test. Some future research projects can be done are the stability of the number of validators. Further research is needed on the number of expert judgments to maximize the index's coefficient. It is better done on both the Aiken formula and the Gregory formula.

## ACKNOWLEDGEMENTS

The authors thank Surabaya State University for its sponsorship and financial support in the Unesa Internal Fund Research Grant number B / 42833 / UN38.III.1 / LK.04.00 / 2023.

## REFERENCES

[1] Ministry of Education and Culture and Technology. Decree of the Minister of Education, Culture, Research, and Technology of the Republic of Indonesia Number 56/M/2022. Ministry of Education and Culture and Technology, 2022b.

[2] A. B. Perdana, S. N. Indah, Y. W. Kusumo, and K. A. Ardhanariswari, "How does the diffusion of kampus merdeka innovation work at the University?" *Proceedings of the International Conference on Engineering, Technology and Social Science (ICONETOS 2020)*, vol. 529, pp. 469–476, Apr. 2021, doi: 10.2991/ASSEHR.K.210421.068.

[3] E. T. Alawiyah and D. Setyorin, "Use of the merdeka teaching platform application in the implementation of the independent curriculum," *International Journal of Information System and Technology (IJISTECH)*, vol. 7, no. 3, pp. 162–167, Oct. 2023.

[4] R. Vebrianto, N. Hermita, D. Irawan, I. M. Mujtahid, and M. Thahir, "Teachers' experiences in sekolah penggerak program: a retrospective case study," *Journal of Education and Learning (EduLearn)*, vol. 18, no. 1, pp. 79–88, Feb. 2024, doi: 10.11591/edulearn.v18i1.20908.

[5] H. Tambunan and T. Naibaho, "Performance of mathematics teachers to build students' high order thinking skills (HOTS)," *Journal of Education and Learning (EduLearn)*, vol. 13, no. 1, pp. 111–117, 2019, doi: 10.11591/edulearn.v13i1.11218.

[6] J. Irvine, "A comparison of revised bloom and Marzano's new taxonomy of learning.," *Res High Educ J*, vol. 33, 2017, Accessed: Aug. 18, 2023. [Online]. Available: https://eric.ed.gov/?id=EJ1161486

[7] D. R. Krathwohl, "A revision of bloom's taxonomy: an overview.," *Theory into practice*, vol. 45, no. 1, pp. 64–65, Jan. 2010, doi: 10.1080/00461520903433562.

[8] F. Zana, C. Sa'dijah, and S. Susiswo, "LOTS to HOTS: How do mathematics teachers improve students' higher-order thinking skills in the class?," *International Journal of Trends in Mathematics Education Research*, vol. 5, no. 3, pp. 251–260, 2022, doi: 10.33122/ijtmer.v5i3.143.

[9] Wardono and S. Mariani, "Increased mathematical literacy and HOTs through realistic learning assisted by e-schoology," *J Phys Conf Ser*, vol. 1567, no. 3, p. 032016, Jun. 2020, doi: 10.1088/1742-6596/1567/3/032016.

[10] Y. M. Sari, A. W. Kohar, Y. I. El Milla, S. Fiangga, and D. S. Rahayu, "Aligning numeracy task design with SDG goals: Nutrition facts as a context for prospective mathematics teachers' problem posing," *Journal on Mathematics Education*, vol. 15, no. 1, pp. 191–206, Nov. 2024, doi: 10.22342/jme.v15i1.pp191-206.

[11] R. P. Yaniawati, "E-learning to improve higher order thinking skills (HOTS) of students," *Journal of Education and Learning (EduLearn)*, vol. 7, no. 2, pp. 109–120, May 2013, doi: 10.11591/edulearn.v7i2.18488.

[12] A. Citrasukmawati, D. Julianingsih, and W. Trisnawaty, "Development of science chess media based on higher order thinking skill (HOTS) to increase the understanding of science concept in students," *J Phys Conf Ser*, vol. 1440, no. 1, p. 012101, Jan. 2020, doi: 10.1088/1742-6596/1440/1/012101.

[13] A. R. Haniah, A. Aman, and R. Setiawan, "Integration of strengthening of character education and higher order thinking skills in history learning," *Journal of Education and Learning (EduLearn)*, vol. 14, no. 2, pp. 183–190, May 2020, doi: 10.11591/edulearn.v14i2.15010.

[14] N. Khusna, S. Bachri, I. K. Astina, D. A. W. Nurhayati, and R. P. Shrestai, "New technologies for project-based empathy learning in merdeka belajar (Freedom to Learn): the use of inaRISK application and biopore technology.," *International Journal of Interactive Mobile Technologies*, vol. 16, no. 22, 2022, doi: 10.3991/ijim.v16i22.36153.

[15] V. L. Lowell and M. Yang, "Authentic learning experiences to improve online instructor's performance and self-efficacy: the design of an online mentoring program," *TechTrends*, vol. 67, no. 1, pp. 112–123, Jan. 2023, doi: 10.1007/S11528-022-00770-5/TABLES/1.

[16] T. Sunarti, Wasis, Supardiyono and M. N. R. Jauhariyah, "Train the skills of making HOTS-based physics questions to physics teachers in Mojokerto," *J Phys Conf Ser*, vol. 1805, no. 1, 2021, doi: 10.1088/1742-6596/1805/1/012027.

[17] Meryansumayeka, Zulkardi, R. Putri and C. Hiltrimartin, "Designing geometrical learning activities assisted with ICT media for supporting students' higher order thinking skills," *Journal on Mathematics Education*, vol. 13, no. 1, pp. 135–148, 2022, doi: 10.22342/jme.v13i1.pp135-148.

[18] I. Y. Zhorova, O. V Kokhanovska, O. M. Khudenko, N. V Osypova, and O. H. Kuzminska, "Teachers' training for the use of digital tools of the formative assessment in the implementation of the concept of the New Ukrainian School," *Educational technology quarterly*, vol. 2022, no. 1, pp. 56–72, 2022, doi: 10.55056/etq.11.

[19] F. N. Husain, "Use of digital assessments how to utilize digital bloom to accommodate online learning and assessments?," *Asian Journal of Education and Training*, vol. 7, no. 1, pp. 2519–5387, 2021, doi: 10.20448/journal.522.2021.71.30.35.

[20] E. Susantini, Isnawati, and Raharjo "HOTS-link mobile learning application: enabling biology pre-service teachers to devise HOTS-based lesson plans," *J Sci Educ Technol*, vol. 31, no. 6, pp. 783–794, Dec. 2022, doi: 10.1007/S10956-022-09993-W.

[21] H. Retnawati, "Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index)," *REiD (Research and Evaluation in Education)*, vol. 2, no. 2, pp. 155–164, 2016, doi: 10.21831/reid.v2i2.11029.

[22] L. R. Aiken, "Three coefficients for analyzing the reliability and validity of ratings," *Educ Psychol Meas*, vol. 45, no. 1, pp. 131–142, 1985, doi: 10.1177/0013164485451012.

[23] A. Sincovich, T. Gregory, C. Zanon, D. D. Santos, J. Lynch, and S. A. Brinkman, "Measuring early childhood development in multiple contexts: The internal factor structure and reliability of the early Human Capability Index in seven low and middle income countries," *BMC Pediatr*, vol. 19, no. 1, Dec. 2019, doi: 10.1186/S12887-019-1852-5.

[24] T. Daan, M. van der Vossen Marianne, W. Anouk, and R. A. Kusurkar, "Validation of the professional identity questionnaire among medical students," *BMC Med Educ*, vol. 21, no. 1, pp. 1–8, Dec. 2021, doi: 10.1186/S12909-021-02704-W/TABLES/5.

[25] J. García-García and C. Dolores-Flores, "Exploring pre-university students' mathematical connections when solving Calculus application problems," *Int J Math Educ Sci Technol*, vol. 52, no. 6, pp. 912–936, 2021, doi: 10.1080/0020739X.2020.1729429.

[26] Y. Sari, B. Kartowagiran, and H. Retnawati, "The characteristics of mathematical reasoning and proof test on indonesian high school students," *J Phys Conf Ser*, vol. 1200, no. 1, p. 12007, 2019, doi: 10.1088/1742-6596/1200/1/012007.

[27] B. Tanujaya and J. Mumu, "Students' misconception of HOTS problems in teaching and learning of mathematics," *J Phys Conf Ser*, vol. 1657, no. 1, p. 12081, 2020, doi: 10.1088/1742-6596/1657/1/012081.

[28] S. Sarwi, S. Saptono, S. Fathonah, and W. Winarto, "Next generation science standards assessment for Java coastal students," *Journal of Education and Learning (EduLearn)*, vol. 17, no. 3, pp. 425–430, Aug. 2023, doi: 10.11591/edulearn.v17i3.20872.

[29] A. Afikah, S. R. D. Astuti, S. Suyanta, J. Jumadi, and E. Rohaeti, "Mobile learning in science education to improve higher-order thinking skills (HOTS) and communication skills: A systematic review," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, 2022, doi: 10.14569/IJACSA.2022.0130782.

[30] L. M. Castro Benavides, J. A. Tamayo Arias, D. Burgos, and A. Martens, "Measuring digital transformation in higher education institutions – content validity instrument," *Applied Computing and Informatics*, 2022, doi: 10.1108/ACI-03-2022-0069/FULL/HTML.

[31] H. N. Hidayati, U. Islam, N. S. Hidayatullah, and W. Rahayu, "Assessment of the validity and reliability of mental health instruments of high school student in Indonesia.," *European Journal of Educational Research*, vol. 10, no. 2, pp. 729–742, 2021, doi: 10.12973/eu-jer.10.2.729.

[32] S. Shankar, C. St-Onge, and M. E. Young, "When I say…response process validity evidence," *Med Educ*, vol. 56, no. 9, pp. 878–880, Sep. 2022, doi: 10.1111/MEDU.14853.

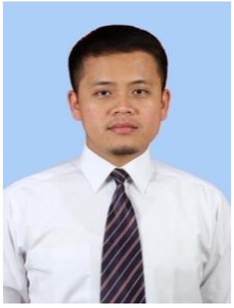
[33] M. Saiful and B. Yusoff, "ABC of content validation and content validity index calculation," *Article in Education in Medicine Journal*, vol. 11, no. 2, pp. 49–54, 2019, doi: 10.21315/eimj2019.11.2.6.


## BIOGRAPHIES OF AUTHORS

**Endang Susantini** is a lecturer at Department of Biology Education, Universitas Negeri Surabaya (UNESA). She obtained bachelor's degree in department of biology education, Institut Keguruan Ilmu Pendidikan Surabaya in 1990 and finished her master's degree program in biology education at Institut Keguruan Ilmu Pendidikan Malang in 1993. She pursued doctoral degree in biology education in 2004 from Universitas Negeri Malang. Her area of interest includes biology education, assessment, metacognitive and higher order thinking skills. She can be contacted at email: endangsusantini@unesa.ac.id.

**Yurizka Melia Sari** is a lecturer at Department of Mathematics Education, Universitas Negeri Surabaya (UNESA). She obtained bachelor's degree in department of mathematics education, Universitas Negeri Surabaya in 2012 and finished her master's degree program in mathematics education at Universitas Negeri Surabaya in 2014. She pursued doctoral degree in educational research and evaluation in 2020 from Universitas Negeri Yogyakarta. Her area of interest includes mathematics education, evaluation, assessment, reasoning, and classroom research. She can be contacted at email: yurizkasari@unesa.ac.id.

**Prima Vidya Asteria** is a lecturer at Department of Indonesian Language Education, Universitas Negeri Surabaya (UNESA). She obtained bachelor's degree in department of Indonesian Language Education, Universitas Negeri Malang in 2011 and finished her master's degree program in Indonesia Language Education at Universitas Negeri Malang in 2013. She pursued doctoral degree in Indonesian Language Education in 2020 from Universitas Negeri Malang. Her area of interest includes Indonesian Language Education, Foreign Language Education, Indonesia as a Second Language Teaching. She can be contacted at email: primaasteria@unesa.ac.id.



**Muhammad Ilyas Marzuqi** is a lecturer at the Department of Social Studies Education, Surabaya State University (UNESA). He obtained his bachelor's degree in Geography Education, Universitas Pendidikan Indonesia, Bandung in 2010 and completed his master's program in Social Studies Education at the State University of Malang in 2016. His areas of interest include Social Studies Education, Geography and History. He can be contacted at email: muhammadmarzuqi@unesa.ac.id.