The Content Validity: Physics Adaptive Test to Measure Higher Order Thinking Skills

Yoga Budi Bhakti[1], Indica Yona Okyranida[2], Rendi Prasetya[3]

{bhaktiyoga.budi@gmail.com[1], indicayona@gmail.com[2], prasetyarendi@gmail.com[3]}

Universitas Indraprasta PGRI, Jl. Raya Tengah No. 80, Jakarta, Indonesia[123]

Abstract. This study is part of the development of a Physics Adaptive Test designed to assess the Higher Order Thinking Skills (HOTS) of eleventh-grade high school students. In this section, the objective is to evaluate the content validity of the test instrument. The developed instrument consists of 20 multiple-choice questions based on the levels of understanding higher-order thinking skills. In addition, this instrument was evaluated both qualitatively and quantitatively by three experts, with the assessment covering the following aspects: 1) Alignment of questions with HOTS indicators, 2) Appropriateness of the stimulus provided in relation to the questions, 3) Clarity in the formulation of questions, 4) Clarity and accuracy in problem descriptions, 5) Logical consistency and homogeneity of answer choices, and 6) The capacity of the instrument items to measure HOTS indicators was evaluated. This study uses a descriptive analysis method to validate the instrument using the Aiken Formula. Content validation data were collected from three physics experts. From the results and analysis, it can be concluded that the 20 items in the Physics Adaptive Test instrument designed to assess HOTS in eleventh-grade physics students are considered valid.

Keywords: Physics Adaptive Test, Test Instrument, Physics

1 Introduction

Higher Order Thinking Skills (HOTS) in physics education encompass students' abilities to analyze complex concepts, evaluate various physical situations, and create creative solutions for unstructured problems [1], [2], [3]. Although the importance of HOTS in physics learning is widely recognized, there is unfortunately no measurement tool or instrument that effectively assesses these skills. Most existing evaluation instruments still focus on measuring basic knowledge and routine problem-solving abilities, without adequately exploring students' critical and creative thinking skills at higher levels [4], [5]. As a result, students' potential to understand and apply physics concepts in new and complex situations often goes unrecognized, limiting opportunities for the optimal development of these skills. The absence of an appropriate instrument to measure students' HOTS in physics underscores the need for the development of

ICIESC 2024, September 17, Medan, Indonesia

DOI 10.4108/eai.17-9-2024.2353081

the Physics Adaptive Test, which can accurately assess and adapt to individual students' abilities to understand and apply physics concepts in complex and dynamic situations.

The development of the Physics Adaptive Test to measure students' Higher Order Thinking Skills (HOTS) arises from the need to enhance the quality of physics learning evaluation in schools. HOTS, which includes analysis, evaluation, and creation abilities, is becoming increasingly important in modern curricula due to the demands of an increasingly complex workforce [6], [7], [8]. However, traditional evaluation approaches often fall short of effectively measuring these skills, particularly in the context of physics, which requires a deep understanding of concepts and the ability to apply them in various situations. Therefore, a more advanced evaluation tool, such as an adaptive test, is needed to measure HOTS more accurately and fairly.

Literature reviews indicate that Computerized Adaptive Testing (CAT) holds great potential in measuring HOTS [9], [10], [11]. Physics adaptive tests work by adjusting the difficulty level of questions based on students' answers, providing each student with a test experience that matches their ability level [12], [13], [14], [15]. This differs from conventional tests that present the same difficulty level to all students, without accounting for individual ability differences. In the context of physics, CAT can provide challenges that are more relevant and appropriate to students' abilities, making it more effective in assessing the extent to which they can apply physics concepts to complex and unstructured problems. Previous research also shows that adaptive tests can reduce student anxiety and increase motivation because they present questions that match the students' ability levels [13], [15].

Thus, the development of the Physics Adaptive Test to measure students' HOTS is not only theoretically relevant but also has significant practical implications. This test can help teachers identify students' weaknesses and strengths more precisely, allowing them to design more targeted instructional strategies. Additionally, adaptive testing can be an effective tool in encouraging students to reach their full potential, as they will face challenging but attainable questions within their capability range. Ultimately, the use of adaptive testing in physics learning evaluation is expected to improve education quality, particularly in preparing students to face future challenges that demand high-level critical and creative thinking [16], [17].
A well-conducted learning assessment process requires a valid instrument to ensure that the assessment results accurately reflect students' abilities and understanding in line with the intended learning objectives. The validity of the assessment instrument is a key factor in determining whether it can accurately measure the intended aspects. A valid instrument provides relevant and reliable data that can be used as a basis for decision-making in the learning process [18]. Without a valid instrument, assessment results may be inaccurate or misleading, preventing the achievement of the actual learning goals. Therefore, the development and validation of assessment instruments are crucial steps in ensuring that the assessment process supports effective and meaningful learning [19], [20], [21].

To ensure that an instrument is considered valid, three common types of validation are used: criterion-related validity, construct validity, and content validity [18]. Criterion-related validity measures the extent to which the instrument's results relate to those from another instrument already proven valid in measuring the same criterion [22]. Construct validity serves to assess whether the instrument actually measures the theoretical concept or construction intended in the preparation of the instrument [23]. Meanwhile, content validity evaluates whether the items within the instrument comprehensively represent all aspects or components of the content being measured [24]. Content validity often involves experts in the relevant field to ensure that each item in the instrument is truly relevant and aligns with the measurement objectives. These three types of validation work together to ensure that the instrument effectively measures its intended construct, ensuring reliable results that are appropriate for the intended assessment objectives.

Based on the description above, the instrument validation stage is a very important step to ensure that the instrument can measure what is designed to be assessed accurately. Hence, this study will evaluate the content validity of the Physics Adaptive Test instrument. The content validation process involves experts in instrumentation, assessment, subject matter, and test development.

2 Method

The research method used for content validation of the Physics Adaptive Test involved a series of systematic steps to ensure that the instrument possesses adequate validity in measuring students' HOTS. The study began with the development of test items based on a literature review and established learning objectives. Once the test items were designed, content validation was conducted by involving a number of experts, including physics lecturers and teachers with expertise in the field.

The experts were tasked with assessing the relevance and suitability of each test item in relation to the measurement objectives, using a Likert scale from 1 (not relevant) to 5 (highly relevant). Their assessments were analyzed with Aiken's V, a statistical method for evaluating content validity. The Aiken's V value was computed for each item to determine how well it aligned with the intended objectives. The experts recorded their evaluations in a structured table provided for this purpose, as shown in Table 1.

Table 1. Question for Evaluation of Instrument

Question	Score				
	1	2	3	4	5
• Questions align with indicators of higher-order thinking skills. • The stimulus presented corresponds appropriately with the questions • Questions are clearly formulated. • Problem descriptions are written clearly and unambiguously. • Answer choices are logical and homogeneous. • Critical thinking skill indicators can be measured through the instrument items.					

In this study, Aiken's V was calculated by considering the number of experts involved, the range of evaluation scores, and the average rating from the experts. Items that received an Aiken's V value above 0.70 were considered to have good content validity and were deemed suitable for inclusion in the instrument. Conversely, items with a value below 0.70 were considered for revision or removal. This approach ensures that the Physics Adaptive Test can accurately and reliably measure HOTS.

3 Result and Discussion

The content validity of the Physics Adaptive Test instrument was established through item assessment by 3 experts. The evaluated aspects included: 1) Alignment of questions with higher-order thinking skills indicators, 2) Appropriateness of the stimulus provided in relation to the questions, 3) Clarity in the formulation of questions, 4) Clarity and accuracy in problem descriptions, 5) Logical consistency and homogeneity of answer choices, and 6) The capacity of the instrument items to evaluate indicators of HOTS.

The validity of the material on the test depends on the logical analysis of the measurement components from the expert, including items, stimulus, test item shape, and sentence structure. It plays a vital role in developing assessment instruments or tests. This process ensures the precision of the measured aspects, captures the complexity of the concept or skill, minimizes measurement bias, enhances the instrument's quality, and improves the objectivity of the research outcomes [25], [26], [27].

In expert content evaluation, key principles for designing multiple-choice tests include: presenting clear content, using language suitable for the test-takers' developmental level, applying standardized language, randomizing the placement of correct answers, ensuring logical consistency among answer options, keeping the length of answer choices relatively similar, and avoiding any hints that reveal the correct answer.

The number of experts participating in this instrument validation aligns with Lynn's recommendation, which suggests involving a minimum of 3 and a maximum of 10 experts. The Physics Adaptive Test instrument, designed to assess students' higher-order thinking skills, provides five answer options for each item. The validation value of each item is calculated based on assessments from 3 experts, with the validation calculation results presented in Table 2.

Table 2. The results of Aiken's V calculation

Item	Judge			s_1	s_2	s_3	$\sum s$	n(c-1)	V	Ket
	Expert 1	Expert 2	Expert 3							
1	4	4	5	3	3	4	32	36	0,89	V
2	5	4	4	4	3	3	28	36	0,78	V
3	4	4	5	3	3	4	32	36	0,89	V
4	3	4	5	2	3	4	26	36	0,72	V
5	4	5	5	3	4	4	32	36	0,89	V
6	4	4	4	3	3	3	26	36	0,72	V
7	5	4	4	4	3	3	33	36	0,92	V
8	5	5	4	4	4	3	30	36	0,83	V
9	4	4	3	3	3	2	30	36	0,83	V
10	5	5	5	4	4	4	31	36	0,86	V
11	5	5	3	4	4	2	30	36	0,83	V
12	5	5	4	4	4	3	33	36	0,92	V
13	4	4	3	3	3	2	30	36	0,83	V

Item	Judge			s_1	s_2	s_3	∑s	n(c-1)	V	Ket
	Expert 1	Expert 2	Expert 3							
14	5	5	5	4	4	4	34	36	0,94	V
15	5	5	3	4	4	2	26	36	0,72	V
16	5	5	4	4	4	3	32	36	0,89	V
17	4	4	3	3	3	2	26	36	0,72	V
18	5	5	5	4	4	4	36	36	1	V
19	5	5	3	4	4	2	29	36	0,81	V
20	5	5	5	4	4	4	33	36	0,92	V

If the Aiken's V value is greater than 0.7, the item is considered valid [28], [29]. This means that the item has met the necessary criteria for content validity, as evaluated by experts. Aiken value above 0.7 indicates strong agreement among experts that the item is appropriately aligned with the intended constructs and accurately measures the aspects it is designed to assess. Therefore, the item can be confidently included in the test or instrument, as it has been validated as a reliable measure of the targeted skills or knowledge. Based on the content validity calculation using Aiken's formula, the results show that the obtained V value is greater than 0.7, indicating that the 20 developed test items are valid. Thus, these 20 items are considered suitable for measuring the predetermined HOTS indicators based on content validity.

Based on the results and analysis of content validation on the Physics Adaptive Test instrument, experts have deemed the instrument valid with some important notes. First, most of the questions are aligned with HOTS indicators such as synthesis, evaluation, and analysis. However, there is still room for improvement in ensuring that all questions truly challenge the test-takers' higher-order thinking abilities. Second, the stimuli presented in the instrument are generally relevant and appropriate to the questions posed. However, some stimuli are suggested to be made more complex to better encourage in-depth analysis by the participants.

Third, in general, the questions are clearly formulated, but some use technical terms or overly complex sentences that could cause confusion. Rewriting these with simpler and more precise language can enhance the clarity of the instrument without reducing the expected depth of thought. Fourth, the problem descriptions in the instrument are clearly and unambiguously written, although there are some items that might be interpreted differently by participants. Therefore, revisions to the ambiguous parts are necessary to ensure uniform understanding among participants.

Fifth, the alternative answers provided are mostly logical and homogeneous, but some items require balancing in terms of difficulty. This is important to ensure that participants truly need to apply critical thinking skills to choose the correct answer. Lastly, the physics adaptive test instruments are quite effective in measuring indicators of students' critical thinking skills in learning physics, although some items need adjustments to be more focused and specific to the indicators being measured. Overall, with the suggested improvements, the instrument is considered valid and is expected to provide more accurate results in measuring critical and HOTS in participants.

Validation carried out by experts serves as proof of validity. The use of expert assessment is the right step to validate the instrument in terms of construction. The validity of the material is carried out through the consensus of experts in the field and is considered valid if the experts believe that the instrument can measure what it should be measured [30], [31], [32]. Experts have the competence to assess the alignment between indicators with the purpose of instrument development, the alignment of indicators with the scope of material or theory, the alignment of instruments with item indicators, and review the correctness of question concepts, answer keys, and the use of language. Therefore, the selection of experts in validating instruments is an important point because it can affect the quality of the instruments developed.

Content validation in the development of adaptive physics test instruments is an important step to produce reliable and accurate test instruments [33], [34]. Without the right content validation process, the instrument risks measuring things that are not in accordance with the goals that have been set. This can lead to inaccurate data collected, incorrect interpretation of results, and decisions made based on those results can be inaccurate. In addition, the invalidity of the instrument can reduce the credibility of the research or evaluation conducted. Expert consensus in the content validation process is measured using the Aiken Formula to assess the quality of the items developed [35]. The content validation process involves a panel of experts who assess each test item based on its relevance and suitability to the concept being measured. Experts use specific scales, such as the Likert scale, to provide a rating, and Aiken's formula is used to calculate the level of agreement between them. Content validation using Aiken's formulas ensures that test items are aligned with measurement objectives and helps identify items that need to be revised or removed to improve the validity and reliability of the instrument in measuring targeted concepts or skills.

4 Conclusion and Suggestion

The conclusion of the study on the development of the Physics Adaptive Test instrument, consisting of 20 items, shows that this instrument has successfully met the validity criteria established by experts. The instrument is effective in measuring HOTS and critical thinking skills in the subjects tested. However, there are still several aspects that require improvement, such as language clarity, the alignment of stimuli with questions, and the balance of answer choices. By revising the items that are still suboptimal, this instrument has the potential to become a more accurate and reliable tool for evaluating adaptive physics abilities in students.

As a recommendation, it is suggested that researchers conduct further trials with a more diverse sample to ensure that this instrument is effective across various educational contexts. Additionally, it is important to continuously update and adjust the instrument items to remain relevant to the evolving curriculum and dynamic learning needs. Thus, this instrument can serve as a stronger evaluation tool and make a significant contribution to the improvement of physics education quality.

Acknowledgement. We sincerely thank the Ministry of Education and Culture of the Republic of Indonesia for their financial support, which has been instrumental in advancing this research project and developing a high-quality evaluation instrument. Our appreciation also goes to the Research and Community Service Institute (LPPM) of Universitas Indraprasta PGRI for their comprehensive support, including the provision of facilities and guidance during the research.

We hope that the findings of this study will have a positive impact on Indonesia's education sector.

References

[1] Apino, E., & Retnawati, H. (2018). Creative problem solving for improving students' higher order thinking skills (HOTS) and characters. In *Character Education for 21st Century Global Citizens* (pp. 249-256). Routledge.

[2] Kwangmuang, P., Jarutkamolpong, S., Sangboonraung, W., & Daungtod, S. (2021). The development of learning innovation to enhance higher order thinking skills for students in Thailand junior high schools. *Heliyon*, *7*(6).

[3] Putranta, H., & Wilujeng, I. (2019, August). Physics learning by PhET simulation-assisted using problem based learning (PBL) model to improve students' critical thinking skills in work and energy chapters in MAN 3 Sleman. In *Asia-Pacific Forum on Science Learning and Teaching* (Vol. 20, No. 1, pp. 1-44). The Education University of Hong Kong, Department of Science and Environmental Studies.

[4] Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series*, *2014*(1), 1-23.

[5] Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking skills and creativity*, *10*, 265-279.

[6] Jaenudin, R., Chotimah, U., Farida, F., & Syarifuddin, S. (2020). Student development zone: higher order thinking skills (hots) in critical thinking orientation. *International Journal of Multicultural and Multireligious Understanding*, *7*(9), 11-19.

[7] Zulmaidah, Z., Suyatna, A., & Rosidin, U. (2020, June). An analysis of need and design of m-learning using scientific approach on electricity material in senior high school to stimulate higher order thinking skills. In *Journal of Physics: Conference Series* (Vol. 1572, No. 1, p. 012005). IOP Publishing.

[8] Zulhelmi, Z., Fauza, N., Syaflita, D., Pratiwi, J., Wijaya, T. T., & Hermita, N. (2023). Development of Learning Media to Improve Students' Higher Order Thinking Skills in Circular Motion Material. *Jurnal Penelitian Pendidikan IPA*, *9*(4), 1734-1740.

[9] Istiyono, E., Dwandaru, W. S. B., Setiawan, R., & Megawati, I. (2020). Developing of Computerized Adaptive Testing to Measure Physics Higher Order Thinking Skills of Senior High School Students and Its Feasibility of Use. *European Journal of Educational Research*, *9*(1), 91-101.

[10] Istiyono, E. (2018, September). IT-based HOTS assessment on physics learning as the 21st century demand at senior high schools: Expectation and reality. In *AIP Conference Proceedings* (Vol. 2014, No. 1). AIP Publishing.

[11] Istiyono, E., Dwandaru, W. S. B., & Faizah, R. (2018). Mapping of physics problem-solving skills of senior high school students using PhysProSS-CAT. *REID (Research and Evaluation in Education)*, *4*(2), 144-154.

[12] Gallego-Durán, F. J., Molina-Carmona, R., & Llorens-Largo, F. (2018). Measuring the difficulty of activities for adaptive learning. *Universal access in the information society*, *17*, 335-348.

[13] Morphew, J. W., Mestre, J. P., Kang, H. A., Chang, H. H., & Fabry, G. (2018). Using computer adaptive testing to assess physics proficiency and improve exam performance in an introductory physics course. *Physical Review Physics Education Research*, *14*(2), 020110.

[14] Istiyono, E. (2019, January). Computer adaptive test as the appropriate model to assess physics achievement in 21st century. In *1st International Conference on Innovation in Education (ICoIE 2018)* (pp. 304-309). Atlantis Press.

[15] Abidin, A. Z., Istiyono, E., Fadilah, N., & Dwandaru, W. S. B. (2019). A Computerized Adaptive Test for Measuring the Physics Critical Thinking Skills. *International Journal of Evaluation and Research in Education*, *8*(3), 376-383.

[16] Batlolona, J. R., Diantoro, M., & Latifah, E. (2019). Creative Thinking Skills Students in Physics on Solid Material Elasticity. *Journal of Turkish Science Education*, *16*(1), 48-61.

[17] Rosen, Y., Stoeffler, K., & Simmering, V. (2020). Imagine: Design for creative thinking, learning, and assessment in schools. *Journal of Intelligence*, *8*(2), 16.

[18] Rahardja, U., Aini, Q., Graha, Y. I., & Lutfiani, N. (2019, December). Validity of test instruments. In *Journal of Physics: Conference Series* (Vol. 1364, No. 1, p. 012050). IOP Publishing.

[19] Balta, N., & Eryilmaz, A. (2020). Development of Modern Physics Achievement Test: Validity and Reliability Study. *European Educational Researcher*, *3*(1), 29-38.

[20] Lina, E. Y., & Desnita, D. (2022). Validity and reliability of critical thinking instruments to measure the effectiveness of context-based physics e-module on wave materials. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *8*(1), 57-64.

[21] Octafia, D., Supriyadi, S., & Sulhadi, S. (2020). Validity and Reliability Content of Physics Problem Solving Test Instrument Based on Local Wisdom. *Journal of Research and Educational Research Evaluation*, *9*(1), 46-51.

[22] Day, E. A., Radosevich, D. J., & Chasteen, C. S. (2003). Construct-and criterion-related validity of four commonly used goal orientation instruments. *Contemporary educational psychology*, *28*(4), 434-464.

[23] Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual review of clinical psychology*, *5*(1), 1-25.

[24] Connell, J., Carlton, J., Grundy, A., Taylor Buck, E., Keetharuth, A. D., Ricketts, T., ... & Brazier, J. (2018). The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Quality of life research*, *27*, 1893-1902.

[25] Jalil, S., Ali, M. S., & Haris, A. (2018, June). Development and validation of science process skills instrument in physics. In *Journal of Physics: Conference Series* (Vol. 1028, No. 1, p. 012203). IOP Publishing.

[26] Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The development of an instrument to measure the higher order thinking skill in physics. *European Journal of Educational Research*, *8*(3), 743-751.

[27] Susiyawati, E., Sudibyo, E., & Sari, D. (2021). Development and Validation of an Instrument. *The International Journal of Assessment and Evaluation*, *28*(2), 1.

[28] Torres-Malca, J. R., Vera-Ponce, V. J., Zuzunaga-Montoya, F. E., Talavera, J. E., La Cruz-Vargas, D., & Jhony, A. (2022). Content validity by expert judgment of an instrument to measure knowledge, attitudes and practices about salt consumption in the peruvian population. *Revista de La Facultad de Medicina Humana*, *22*(2), 9.

[29] Penfield, R. D., & Giacobbi, Jr, P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in physical education and exercise science*, *8*(4), 213-225.

[30] Wati, M., & Misbah, M. (2021). The content validity of the assessment instrument on the characters of wasaka in wetland environment physics learning. In *Journal of Physics: Conference Series* (Vol. 1760, No. 1, p. 012016). IOP Publishing.

[31] Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2024). The Content Validity: Multiple Choices Complex Physics Instrument to Measure Critical Thinking Skills. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *9*(1), 56-64.

[32] Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2023, September). Development of an assessment instrument for critical thinking skills in Physics: a systematic literature review. In *Journal of Physics: Conference Series* (Vol. 2596, No. 1, p. 012067). IOP Publishing.

[33] Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological assessment*, *31*(12), 1412.

[34] Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in social and administrative pharmacy*, *15*(2), 214-221.

[35] Kholis, N., Kartowagiran, B., & Mardapi, D. (2020). Development and Validation of an Instrument to Measure a Performance of Vocational High School. *European Journal of Educational Research*, *9*(3), 955-966.