



## Original Article

### Analysis of the Validity of the Content of Daily Test Instruments in Science Subjects in Grade V of Kromong State Elementary School

Parmi<sup>1✉</sup>, Sa'idatud Daroini<sup>2</sup>, Benny Chandra Nahera<sup>3</sup>, Muhammad Na'am<sup>4</sup>, Nataniel Ampung<sup>5</sup>, Nensy Megawati Simanjuntak<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Universitas Dr. Soetomo Surabaya

Correspondence Author: ny.parmi1981@gmail.com ✉

#### Abstract:

This study aims to analyze the content validity of the daily test instrument for the Natural and Social Sciences (IPAS) subject for fifth graders at Kromong Public Elementary School. Content validity is a fundamental aspect in the development of evaluation instruments, as it ensures that each item truly represents the basic competencies and learning indicators to be measured. This study used a quantitative descriptive approach involving three validators consisting of an IPAS material expert, an educational evaluation expert, and a senior teacher practitioner. The validated instrument consisted of eight descriptive questions on the theme "Living Things and Their Environment". The assessment was carried out using a four-point Likert scale and analyzed using the Aiken's V index to measure the level of agreement between validators regarding the suitability of the items. The analysis results showed that six of the eight items were in the high category with V values between 0.89 and 1.00, while the other two items obtained the medium category with V values of 0.78. The overall average Aiken's V value was 0.89, which indicates that the instrument has high content validity and is suitable for use in learning evaluation. These findings confirm that the cross-competency expert review-based validation process plays a significant role in ensuring the instrument's quality. However, some items require revision in terms of construction and language to improve clarity and understanding for elementary school students. Overall, the instrument can be used as a daily assessment tool with minor improvements to certain items.

**Keywords:** Aiken's V, daily tests, elementary school, evaluation instrument

#### Introduction

Evaluation of learning outcomes is a key component of the education system, serving to assess the extent to which learning objectives have been achieved. One form of

Submitted	: 10 January 2026
Revised	: 21 January 2026
Acceptance	: 30 January 2026
Publish Online	: 31 January 2026

evaluation routinely conducted in elementary schools is daily tests, which serve as a continuous measure of student learning progress ([Ramadhan et al., 2023](#)). Through daily tests, teachers can identify students' level of understanding of the material taught and also serve as a basis for decision-making in subsequent learning processes ([Fuadiy, 2021](#)). However, the reliability of test results is highly dependent on the quality of the instruments used. If the instruments used do not meet validity and reliability standards, test results can provide an inaccurate picture of student abilities and potentially lead to errors in assessment ([Sarwanto et al., 2021](#)).

Content validity is a crucial aspect in determining the quality of an evaluation instrument. Content validity indicates the extent to which the test items truly represent the content of the material that aligns with the core competencies and learning indicators ([Almanasreh et al., 2019](#)). Therefore, content validity ensures that each test item is relevant and represents the domain of knowledge being measured, so that evaluation results can comprehensively describe student abilities, not merely by chance or other irrelevant aspects ([Shuttleworth254.5k, 2020](#)). The content validity assessment process typically involves a panel of experts who assess the suitability and appropriateness of each test item for the learning objectives, ensuring that the instrument used can be trusted to objectively measure competency achievement ([Lestari & Rispatiningsih, 2023](#)).

The content validation process is not simply an administrative activity, but rather a systematic procedure involving several important stages. This process generally begins with curriculum analysis and the determination of relevant material domains, followed by the development of a structured grid or blueprint to ensure the representation of each aspect of the competency being measured ([Pawade et al., 2022](#)). Next, the developed instrument undergoes in-depth review by a panel of experts selected based on their expertise and experience in the relevant field. The experts assess the relevance, clarity, and appropriateness of each item using qualitative and quantitative methods, such as the Content Validity Index (CVI) or Likert scale assessments ([Suhaini et al., 2021](#)).

In practice, content validity is often combined with a quantitative approach to increase the objectivity of expert assessments. One commonly used method is the calculation of the Aiken's V coefficient and the Content Validity Ratio (CVR) developed by Lawshe (1975) ([Jeldres et al., 2023](#)). The value of this coefficient indicates the level of agreement among experts regarding the relevance of a test item; the higher the coefficient value, the more valid the instrument developed ([Tajuddin et al., 2025](#)). This quantitative approach is widely used in educational instrument development research because it provides a strong statistical basis for assessing qualitative validation results, allowing the item selection and revision process to be carried out more objectively and measurably ([Chong et al., 2021](#)).

Various previous studies have also emphasized the importance of content validity in ensuring the accuracy of evaluation instruments. A study by [Muliana et al. \(2020\)](#) found that of the 39 items developed, only 26 met content validity criteria with an Aiken's V coefficient above 0.87, while the rest had to be eliminated or revised. [Setiawan et al. \(2024\)](#) reported that the learning evaluation instrument they developed had an average Aiken's V value of 0.75, which is considered high, but several items still required refinement based on expert input. Another study by [Kania et al. \(2024\)](#) showed that the mathematical problem-solving ability assessment instrument had a content validity value between 0.817 and 0.884, indicating a very good level of validity. Meanwhile, [Anggara and Abdillah \(2023\)](#) also reported that all items in the literacy assessment

instrument tested had an average Content Validity Index (CVI) value of 0.96, thus being declared content valid. These findings show that the content validation process is an essential step in ensuring the quality and accuracy of test instruments before they are used in learning evaluations.

However, the reality on the ground shows that not all schools systematically implement content validation in developing evaluation instruments. Many teachers still develop questions based on personal experience without referring to proper validation procedures, such as using blueprints or expert reviews. As a result, most daily tests or teacher-made tests do not fully reflect the core competencies outlined in the curriculum, resulting in a mismatch between the material tested and the learning objectives ([Tiana & Maruf, 2024](#)).

At Kromong Public Elementary School, this phenomenon is also evident in learning evaluation practices, particularly in the fifth-grade Natural and Social Sciences (IPAS) subject. IPAS is an integrative subject that combines scientific, social, and environmental concepts. The complexity of this material demands the development of valid evaluation instruments to holistically measure student abilities. However, to date, there has been no in-depth study assessing the extent to which the content validity of the daily IPAS test instruments at the school has been met.

The Merdeka Curriculum, which has begun to be implemented in many elementary schools, emphasizes the importance of authentic assessments that assess not only learning outcomes but also students' thinking processes and conceptual understanding. Therefore, the content validity of daily test instruments is increasingly crucial to ensure the alignment of questions with learning outcomes. Through systematic content validation, teachers can develop evaluation instruments that are not only accurate but also relevant to the needs of competency-oriented 21st-century learning. This study aims to analyze the content validity of daily test instruments in fifth-grade science at Kromong Public Elementary School.

## Methods

This study uses a quantitative descriptive approach, which aims to objectively analyze and describe the content validity level of daily test instruments in Natural and Social Sciences (IPAS). This approach was chosen because it provides a factual picture of the instrument's quality based on expert assessments, which are then processed using numerical data and analyzed statistically. Using a quantitative descriptive method, this study does not focus on hypothesis testing, but rather on interpreting the validation results carried out by experts to determine the extent to which the test items are in accordance with basic competencies, learning indicators, and the rules for writing educational evaluation instruments.

The research subjects consisted of three validators selected based on their expertise and experience in the field of elementary education and the development of evaluation instruments. The three validators included: (1) an IPAs material expert, namely an elementary school supervisor who assessed the suitability between the test items and the teaching materials; (2) an educational evaluation expert, namely a lecturer who assessed the construction of the questions from the aspects of language, readability, and rules of composition; and (3) a field practitioner, namely the head of the KKKS who acted as a senior class teacher and provided practical views regarding the suitability of the instrument in the classroom. The object of the research was the daily IPAs test instrument for grade V of Kromong State Elementary School with the theme "Living

Things and Their Environment", which consisted of a number of multiple-choice questions developed based on the applicable curriculum.

This research procedure was carried out in three main stages, namely instrument development, expert validation, and data analysis. In the first stage, the researcher conducted a curriculum analysis to identify basic competencies, compile a question grid, and develop instrument items according to learning indicators. The second stage, namely expert validation, was carried out by providing assessment sheets to three validators. Each expert assessed the suitability of the content, language clarity, and question construction structure using a four-point Likert scale (1 = not appropriate, 2 = less appropriate, 3 = appropriate, 4 = very appropriate). The final stage was data analysis, namely calculating the level of content validity using the Aiken's V index. The Aiken's V value was used to measure the degree of agreement between validators regarding the relevance of each question item to basic competencies and learning indicators.

Validation data were analyzed quantitatively descriptively using Aiken's V formula. The results of the V coefficient calculation were interpreted to determine the validity level of each item. The interpretation criteria refer to Retnawati (2016), namely: a V value  $< 0.40$  is categorized as low,  $0.40 \leq V < 0.80$  is categorized as medium, and  $V \geq 0.80$  is categorized as high. Items with a V value  $\geq 0.80$  are declared valid and suitable for use, while items with medium or low values are recommended for revision. The final data are presented in the form of tables and descriptive descriptions to describe the strengths and weaknesses of the daily science test instrument as a whole. Through this analysis, the study is expected to provide an objective picture of the quality of the instrument's content validity and become a reference for teachers in developing more accurate and quality evaluation tools.

## Results

Content validity analysis was conducted on 8 descriptive questions on the Natural Sciences subject, the theme of Living Things and Their Environment, for fifth-grade students of Kromong State Elementary School. Validation was carried out by three expert validators. The assessment used a 4-point Likert scale and was analyzed using the Aiken's V formula to measure the level of suitability of the question items with the learning indicators.

Table 1. Validation results by three validators

Question	Validator 1	Validator 2	Validator 3
The relationship between living things and their environment	4	4	3
Adaptation of living things and their environment	4	4	4
The purpose of living things to adapt	3	4	3
Characteristics of living things	4	3	4
The importance of sunlight for plants	4	4	3
Definition of ecosystem	4	3	3
The impact of water pollution	4	4	4
Benefits of maintaining environmental ecosystems	4	4	3

Table 2. Aiken's V index per item

Question	V	Category
The relationship between living things and their environment	0,89	High
Adaptation of living things and their environment	1,00	High
Purposes and examples of living things adapting	0,78	Medium
Characteristics of living things	0,89	High
The importance of sunlight for plants	0,89	High
Definition of ecosystems	0,78	Medium
Impacts of water pollution	1,00	High
Benefits of maintaining environmental ecosystems	0,89	Medium

The overall average Aiken's V score was 0.89, categorized as high content validity. Validation results indicate that the majority of test items met the criteria for content suitability with learning indicators. Test items 2 and 7 received perfect scores ( $V = 1.00$ ), indicating full agreement among validators that both items were highly relevant. Factors contributing to this high score include clarity of wording, alignment with core competencies, and appropriate levels of difficulty.

Two questions, numbers 3 and 6, received a moderate rating ( $V = 0.78$ ). Based on validator feedback, question 3 combined two concepts—geographic location and its impact on the economy—which were deemed too complex to combine. Separating the questions or rephrasing the sentence is recommended to help students understand the question's requirements more clearly. In question 6, the term "definition of ecosystem," although common in education, is potentially difficult for students to grasp. Simplifying the term to "understanding ecosystem" could improve understanding.

The results of the study indicate that the fifth-grade science daily test instrument at Kromong Public Elementary School has met high content validity based on the Aiken's V index analysis. The average V value of 0.89 confirms that most of the test items have strong relevance to the established learning indicators. This finding emphasizes the importance of the validation process as a systematic step to ensure the alignment between the evaluation instrument and the learning objectives. Instrument validation involving a panel of experts has proven to be the main foundation in ensuring the relevance and accuracy of the test items to the expected competencies (Suhaini et al., 2021). Through assessment and input from experts, each instrument item can be evaluated in depth so that only items that truly align with the curriculum and learning needs are retained (Jamali et al., 2024).

The results of the study indicate that the fifth-grade science daily test instrument at Kromong Public Elementary School has met high content validity based on the Aiken's V index analysis. The average V value of 0.89 confirms that most of the test items have strong relevance to the established learning indicators. This finding emphasizes the importance of the validation process as a systematic step to ensure the alignment between the evaluation instrument and the learning objectives. Instrument validation involving a panel of experts has proven to be the main foundation in ensuring the relevance and accuracy of the test items to the expected competencies (Almanasreh et al., 2019). Therefore, the high Aiken's V value for most of the items, as found in various

studies with a value range above 0.75 to 0.95, indicates that the instrument is on track in accordance with the theoretical standards for educational test development ([Suciati et al., 2020](#)).

The involvement of three validators with diverse backgrounds—a content expert, an educational evaluation expert, and a senior teacher—contributed significantly to the instrument's quality. This cross-functional approach to instrument validation proved crucial because it involved diverse expert perspectives, which increased assessment accuracy and enriched the item revision process. As emphasized by [Syafitri et al. \(2023\)](#), input from eight validators, consisting of teachers and lecturers, complemented each other in assessing aspects of the material, construction, and language, resulting in a comprehensive picture of the instrument's strengths and weaknesses. This process allowed each validator to provide a distinct perspective based on their expertise, resulting in more comprehensive revisions and a more representative instrument suitable for use in learning evaluations ([Suciati et al., 2020](#)).

The high validity scores on certain items such as “Adaptation of living things and their environment” and “Impact of water pollution”, both of which received a perfect score ( $V = 1.00$ ), indicate full agreement between validators regarding the relevance of the concept to the learning outcomes of grade V. In addition to clear wording, these questions are considered capable of reflecting the essential concepts in the theme of Living Things and Their Environment, as mandated in the Natural Sciences curriculum at the elementary school level.

However, there are two questions that fall into the moderate category, namely "The Purpose of Living Things to Adapt" and "Definition of an Ecosystem," each with a  $V$  value of 0.78. Based on the validator's notes, both questions have construction aspects that need to be improved. In question number 3, the problem arises from combining two concepts in one question, which is considered to be confusing for students. In question number 6, the problem lies in the term "definition," which is considered unfamiliar to fifth-grade students, thus potentially reducing the readability of the question.

Linguistic and item construction issues such as those found in this moderate category item have also been frequently encountered in previous research. Research by [Lo et al. \(2021\)](#) showed that linguistic demands, particularly in the form of complex or productive sentences, significantly impacted students' test performance, even when the material being tested was relevant. [Zdorova et al. \(2025\)](#) also found that although two texts in the LexiMetr assessment tool had similar linguistic complexity, students' subjective difficulty levels and reading speeds differed, indicating that language structure and clarity significantly influence assessment results. These findings confirm that linguistic aspects, such as simplicity of sentence structure and clarity of terminology, are crucial for the content validity of an instrument and, therefore, should be a primary concern when developing items for elementary school-aged students.

The validation process conducted in this study aligns with the systematic steps for content validity as described in various literature, starting with the development of instrument specifications, review by a panel of experts, and then quantitative analysis using the Aiken's  $V$  index ([Kania et al., 2024](#)). Each instrument item was evaluated by experts to assess its suitability with the indicators and measurement objectives. The results were then statistically analyzed to determine the level of content validity ([Ulya et al., 2024](#)).

## Conclusion

The daily science test instrument for grade V of Kromong State Elementary School has been proven to have a high level of validity with an average value of 0.89. The majority of questions are considered relevant, representative, and in accordance with learning indicators, making it suitable for use as an evaluation tool. Two questions in the moderate category require minor revisions, especially in the construction and language aspects, to improve clarity and understanding for students. Overall, this instrument has met the content validity standards and can be an appropriate reference in assessing student competency achievement in the material "Living Things and Their Environment".

## Suggestion

It is recommended that teachers and instrument developers revise items in the moderate category, particularly those related to simplifying wording and separating overly complex concepts. The validation process should be carried out continuously, involving a panel of experts from various backgrounds, to ensure the instrument's quality is maintained. Furthermore, in further instrument development, teachers need to integrate authentic assessment principles aligned with the Independent Curriculum so that the instrument measures not only factual knowledge but also students' critical thinking skills, conceptual understanding, and scientific process skills.

## References

- Almanasreh, E., Moles, R., & Chen, T. (2019). Evaluation of Methods Used for Estimating Content Validity. *Research in Social and Administrative Pharmacy*, 15, 214. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Anggara, D., & Abdillah, C. (2023). Content Validity Analysis of Literacy Assessment Instruments. *Jurnal Cakrawala Pendidikan*. <https://doi.org/10.21831/cp.v42i2.55900>
- Chong, J., Mokshein, S. E., & Mustapha, R. (2021). A Content Validity Study for Vocational Teachers' Assessment Literacy Instrument (VoTAL). *International Journal of Academic Research in Business and Social Sciences*. <https://doi.org/10.6007/ijarbss/v11-i4/9547>
- Fuadiy, M. (2021). Evaluasi Pembelajaran Sebagai Sebuah Studi Literatur. *DIMAR: Jurnal Pendidikan Islam*. <https://doi.org/10.58577/dimar.v3i1.83>
- Jamali, A. M., Chandio, S., & Maree, M. (2024). Investigating The Content Validity of An Iot-Acceptance Survey Instrument in Higher Education: A Study in The Context of Sindh, Pakistan. *The Asian Bulletin of Big Data Management*. <https://doi.org/10.62019/abbdm.v3i2.78>
- Jeldres, M. R., Costa, E. D., & Nadim, T. F. (2023). A Review of Lawshe's Method for Calculating Content Validity in the Social Sciences. *Frontiers in Education*. <https://doi.org/10.3389/educ.2023.1271335>
- Kania, N., Kusumah, Y., Dahlan, J. A., Nurlaelah, E., Gürbüz, F., & Bonyah, E. (2024). Constructing and Providing Content Validity Evidence Through the Aiken's V Index Based on the Experts' Judgments of the Instrument to Measure Mathematical Problem-Solving Skills. *REID (Research and Evaluation in Education)*. <https://doi.org/10.21831/reid.v10i1.71032>
- Lestari, W., & Rispatiningsih, D. M. (2023). Validity of the TPACK Instrument Content for Prospective Chemistry Teachers using the Rasch Model Application. *JIMAD : Jurnal Ilmiah Mutiara Pendidikan*. <https://doi.org/10.61404/jimad.v1i2.87>
- Lo, Y., Fung, D., & Qiu, X. (2021). Assessing Content Knowledge Through L2: Mediating Role of Language of Testing on Students' Performance. *Journal of*

- Multilingual and Multicultural Development*, 44, 1013-1028. <https://doi.org/10.1080/01434632.2020.1854274>
- Muliana, N., Pada, A., & Nurmaliah, C. (2020). Content validity of conation assessment. *Journal of Physics: Conference Series*, 1460. <https://doi.org/10.1088/1742-6596/1460/1/012057>
- Pawade, Y., Chalak, A., & Pawade, D. (2022). Blueprints in Biochemistry: Systematic Assessment in Undergraduate Medical Education. *Research and Development in Medical Education*. <https://doi.org/10.34172/rdme.2022.022>
- Ramadhan, W., Malahati, F., Romadhon, K., & Ramadhan, S. (2023). Analysis of Multiple Choice Questions in Elementary School Daily Assessments. *Tarbiyah Wa Ta'lim: Jurnal Penelitian Pendidikan dan Pembelajaran*. <https://doi.org/10.21093/twt.v10i2.6155>
- Sarwanto, S., Fajari, L., & Chumdari, C. (2021). Open-Ended Questions to Assess Critical-Thinking Skills in Indonesian Elementary School. *International Journal of Instruction*, 14, 615-630. <https://doi.org/10.29333/iji.2021.14137a>
- Setiawan, R., Wagiran, W., & Alsamiri, Y. (2024). Construction of an Instrument for Evaluating the Teaching Process in Higher Education: Content and Construct Validity. *REID (Research and Evaluation in Education)*. <https://doi.org/10.21831/reid.v10i1.63483>
- Shuttleworth254.5k, M. (2020). Content Validity. *Definitions*. <https://doi.org/10.4135/9781412952644.n104>
- Suciati, Munadi, S., Sugiman, & Febriyanti, W. D. R. (2020). Design and Validation of Mathematical Literacy Instruments for Assessment for Learning in Indonesia. *European Journal of Educational Research*, 9, 865-875. <https://doi.org/10.12973/eu-jer.9.2.865>
- Suhaini, M., Ahmad, A., & Mohd, N. (2021). Assessments on Vocational Knowledge and Skills: A Content Validity Analysis. *European Journal of Educational Research*. <https://doi.org/10.12973/eu-jer.10.3.1529>
- Syafitri, A., Anggraini, D., Parinduri, W. M., Rambe, T. R., & Rambe, N. (2023). Content Validity Analysis Of Final Semester Assessment Instruments For Science Subjects In Elementary School. *SCHOOL EDUCATION JOURNAL PGSD FIP UNIMED*. <https://doi.org/10.24114/sejpgsd.v13i4.54309>
- Tajuddin, N. I. I., Abas, U., Aziz, K. A., Nor, R., Izni, N., Sudin, M. N., Anim, N. A. H. M., & Noor, N. (2025). Content Validity Assessment Using Aiken's V: Knowledge Integration Model for Blockchain in Higher Learning Institutions. *International Journal of Advanced Computer Science and Applications*. <https://doi.org/10.14569/ijacsa.2025.0160659>
- Tiana, E. A., & Maruf, N. (2024). Aligning Assessment and Curriculum: A Study of Summative Tests in Educational Settings. *JET ADI BUANA*. <https://doi.org/10.36456/jet.v9.n01.2024.8884>
- Ulya, H., Sugiman, S., & Rosnawati, R. (2024). Design and Content Validity of Mathematics Creative Problem-Solving Ability Instrument for Junior High School Students. *Eurasia Journal of Mathematics, Science and Technology Education*. <https://doi.org/10.29333/ejmste/14661>
- Zdorova, N., Staroverova, V., Lopukhina, A., Dzhonboboeva, D., Pareskina, E., Shestakova, E., Gmyrina, V., Dmitrova, E. P., & Dragoy, O. (2025). Reading Assessment Tool "LexiMetr" for Russian-Speaking Primary School Students. *Voprosy obrazovaniya / Educational Studies Moscow*. <https://doi.org/10.17323/vo-2025-21693>