



Content Validity Analysis of Multitier Diagnostic Test Instruments for Measuring Junior High School Disaster Literacy

Apriyani¹, Henny Johan^{1*}, Rahmawati², Muchammad Farid³, Sutarno¹, Aprina Defianti¹

¹Science Education, Universitas Bengkulu, Bengkulu, Indonesia

²Master of Elementary Education, Universitas Muhammadiyah Makassar, Makassar, Indonesia

³Geophysics, Universitas Bengkulu, Bengkulu, Indonesia

✉ * hennyjohan@unib.ac.id

Article Info	Abstract
Keywords: <i>four-tier diagnostic tests; disaster literacy; content validity; Content Validity Ratio; Content Validity Index</i>	Background: Disaster literacy is an essential competency for students in disaster-prone areas such as Bengkulu Province. However, conventional school assessments cannot adequately measure students' conceptual understanding, misconceptions, and disaster literacy. Moreover, no validated four-tier diagnostic instrument is available for vibration and wave topics. Methods: This study used the ADDIE R&D model, with research limited to the development phase. Five experts validated the 32-item instrument. Two university lecturers and three junior high school teachers served as experts. The Content Validity Ratio (CVR) and Content Validity Index (CVI) were used to assess content validity. Results: The CVR values for all 32 test items ranged from 0.81 to 1.00. Based on Lawshe's criteria, 18 items met the minimum CVR requirement, while the remaining 14 were revised per the validators' recommendations. The Content Validity Index (CVI) was 1.00 for the knowledge and attitude aspects and 0.96 for the skills aspect, yielding an overall CVI of 0.972, indicating excellent content validity. The revised instrument was then deemed ready for implementation with junior high school students to collect empirical response data for subsequent validity and reliability analyses. Conclusion: The developed disaster-literacy-based four-tier diagnostic test demonstrated excellent content validity following expert validation and revision. This study contributes a validated four-tier instrument for assessing disaster literacy and students' conceptual understanding in vibration and wave topics. Despite the challenges of developing coherent four-tier items, the instrument was considered suitable for empirical testing with junior high school students.
Received: Date 24/07/2026 Revised: 05/09/2026 Accepted: Date 07/09/2026	

INTRODUCTION

Indonesia is the country with the second-highest risk of natural disasters in the world. According to Hilft (2023) Indonesia ranks second after the Philippines in terms of disaster

vulnerability. The National Disaster Management Agency (BNPB) recorded 4.852 natural disaster events throughout 2023, most of which were floods, extreme weather events, forest and land fires, landslides, droughts, earthquakes, and volcanic eruptions. The high frequency of disasters is largely attributed to Indonesia's geographical and geological characteristics. The country is located at the convergence of the Eurasian, Indo-Australian, and Philippine tectonic plates and has a tropical climate, making it highly susceptible to various types of natural disasters (Halik & Septiana, 2022).

Bengkulu Province is one of the regions in Indonesia that has a high level of disaster vulnerability. According to BNPB (2019), Bengkulu is prone to various types of disasters, including earthquakes, tsunamis, floods, landslides, droughts, and tornadoes. Geologically, Bengkulu is situated in the subduction zone between the Indo-Australian and Eurasian tectonic plates. Geographically, Bengkulu City is located along the western coast of Sumatra island, directly facing the Indian Ocean. These geological and geographical conditions place Bengkulu at high risk of disasters, particularly earthquakes and tsunamis (Safitra et al. 2023; Susmita et al. 2023). As a result, it requires the community, especially students, to have good knowledge and skills related to disaster literacy to minimize disaster risk when the event occurs.

The high risk of disasters in Indonesia has prompted the government to develop various disaster risk reduction programs through the education sector. To reduce disaster risk through the education sector, the Indonesian Government, through the National Disaster Management Agency (BNPB), has implemented the Disaster Safe Education Unit (SPAB) Program as a national initiative to integrate disaster risk reduction into school education and strengthen students' disaster literacy and preparedness (Anisah, 2019). However, the successful implementation of SPAB also requires appropriate assessment instruments capable of measuring students' disaster literacy accurately, enabling teachers to evaluate whether the objectives of the program have been achieved. The program emphasizes that disaster education focuses not only on mastering concepts, but also on building students' knowledge, skills, and preparedness in dealing with disasters (Suleman, 2024; Handayani et al. 2024). Although the SPAB program has become one of the government's efforts to improve disaster literacy in schools, its implementation in various regions still faces challenges, especially in the integration of disaster materials into the learning and assessment of students' disaster literacy skills (Tahmidaten & Krismanto, 2019; Sudrajat & Wibowo, 2016).

Schools play a crucial role in fostering students' disaster preparedness through disaster education. Disaster education not only provides knowledge about the types and causes of disasters but also equips students with the ability to understand, interpret, and utilize disaster-related information to support mitigation, preparedness, emergency response, and post-disaster recovery efforts (Amelia et al. 2023). Adequate disaster literacy contributes significantly to improving students' preparedness for disaster situations (Putri et al. 2025). However, previous studies have shown that disaster literacy among junior high school students remains inadequate. For example, Berlianawati et al. (2025) reported that students demonstrated limited understanding of disaster mitigation, indicating insufficient knowledge and preparedness to respond effectively to disaster situations. Consistent with these findings, the needs assessment conducted in this study involving 78 eighth-grade students in Bengkulu City revealed that the integration of disaster education into science learning (66%) and students' preparedness (74%) were only at a moderate level. Interviews with science teachers further indicated that disaster-related topics had not been systematically integrated into science instruction and that classroom assessments were still dominated by conventional tests, which were unable to comprehensively identify students' conceptual understanding. Moreover, teachers had never implemented multitier

diagnostic assessments, particularly four-tier diagnostic tests, to evaluate students' disaster literacy.

This finding is consistent with previous studies showing that disaster literacy is primarily assessed using multiple-choice and essay-based assessments, which mainly measure students' final answers rather than their underlying reasoning and confidence. Junianto & Hendriani (2024) found that the disaster preparedness tests used in schools were still multiple-choice objective tests, whereas Islamic *et al.* (2025) measured disaster literacy using essay questions and questionnaires. Although these assessment methods can provide information about students' learning outcomes, they cannot reveal the reasoning behind students' answers or their level of confidence. Consequently, students' conceptual understanding cannot be accurately identified.

Numerous research efforts have shown that four-level diagnostic evaluations are superior to traditional methods in detecting students' conceptual knowledge and misunderstandings. Four-tier diagnostic tests are capable of distinguishing between students who correctly understand concepts and those who hold misconceptions (Rahmawati *et al.* 2023). Similarly, Fiskawarni *et al.* (2024) developed a four-tier instrument to measure energy literacy, Priasmika & Sholichah (2022) developed a four-tier diagnostic instrument to identify misconceptions related to salt hydrolysis. These findings indicate that four-tier diagnostic instruments have been widely applied to various science topics; however, they have not yet been developed to measure students' disaster literacy. Therefore, this approach has strong potential to be adapted for assessing disaster literacy, as it can evaluate not only students' answers but also the reasoning and confidence underlying their responses. However, to the best of the authors' knowledge, no four-tier diagnostic instrument has yet been specifically developed to measure disaster literacy among junior high school students.

Johan *et al.* (2021) stated that disaster-related topics are closely associated with science education and are appropriate for integration into science learning, particularly in disaster-prone areas. However, no previous study has developed a validated four-tier diagnostic instrument to assess disaster literacy among junior high school students in the context of vibration and wave topics, despite the close relationship of these topics to earthquake and tsunami phenomena. Consequently, the development of such an instrument requires rigorous content validation before implementation to ensure that it accurately measures the intended construct. Expert validation enables specialists to evaluate the relevance, clarity, and representativeness of the instrument, identify weaknesses, and provide recommendations for revision (Nainggolan *et al.* 2026). In instrument development, this process is referred to as content validity, which determines whether the test items adequately represent the intended concepts, indicators, and measurement objectives (Lawshe, 1975). Content validity is commonly evaluated using the Content Validity Ratio (CVR) and Content Validity Index (CVI), which quantify the degree of expert agreement regarding the relevance of each item (Polit & Beck, 2006). Therefore, CVR and CVI analyses constitute essential procedures for ensuring that a four-tier diagnostic instrument possesses adequate content validity before empirical testing.

Several studies have examined the content validity of four-tier diagnostic test instruments. Uswatun & Mubarak (2024) developed a *Four-tier* diagnostic test on the topics of the mole concept and stoichiometry, which was evaluated by four experts using the CVI method and achieved a very high validity category. Similarly, Febriana & Margereta (2026) developed a four-tier diagnostic test on cell biology and reported high content validity based on Aiken's *V* (average = 0.88), accompanied by good reliability ($KR-20 = 0.78$). Their findings also demonstrated that the instrument effectively distinguished students who understood the concepts, held misconceptions, or lacked conceptual understanding based on the combination of answers, reasoning, and confidence levels. These findings emphasize that expert validation is a crucial stage in ensuring the content validity of an instrument before it is administered to users (Nainggolan *et al.* 2026).

Despite these developments, many previous studies have focused largely on four-tier diagnostic instruments for science subjects. This study specifically examines the content validity of a multi-tier diagnostic test instrument designed to evaluate disaster literacy among junior high school students. This study aims to establish the content validity of a multi-tier diagnostic test instrument to measure disaster literacy in junior high school students through expert assessment using the Content Validity Ratio (CVR) and Content Validity Index (CVI) approaches. Therefore, this study aims to analyze the content validity of multi-tier diagnostic test instruments using expert assessment with the Content Validity Ratio (CVR) and Content Validity Index (CVI) approaches as the basis for instrument feasibility before conducting empirical testing.

METHODS

This study used the R&D method with the ADDIE model by Branch (2009). The instrument development process consisted of five stages: Analysis, Design, Development, Implementation, and Evaluation. During the Analysis stage, a needs assessment was conducted through a literature review, student questionnaires, and interviews with science teachers to identify the need for a disaster literacy diagnostic instrument. The Design stage involved developing the assessment indicators, test blueprint, and multitier diagnostic test items based on disaster literacy in the context of vibration and wave topics. During the Development stage, the instrument was constructed and evaluated by five expert validators using the Content Validity Ratio (CVR) and Content Validity Index (CVI). The Implementation stage involved administering the instrument to junior high school students, while the Evaluation stage was conducted formatively based on expert feedback and trial results to refine the instrument. Although all five stages of the ADDIE model were implemented during the instrument development process, this article focuses on the Analysis, Design, and Development stages. The Implementation and Evaluation stages are described only to provide the overall development framework, whereas the findings reported in this article primarily emphasize the expert validation conducted during the Development stage.

The instrument was developed from the findings of the needs assessment and a review of the literature on disaster literacy and the concepts of vibration and waves. A total of 32 multitier diagnostic test items were constructed by integrating the concepts of vibration, waves, earthquakes, and tsunamis into authentic disaster contexts. The content validity of the instrument was evaluated using the Content Validity Ratio (CVR) and the Content Validity Index (CVI). The data source in this study consisted of five expert validators, including two university lecturers and three junior high school science teachers from schools in Bengkulu City. The validators were selected based on their expertise in science education and educational instrument development. The science teachers were purposively selected from schools located in disaster-prone areas to ensure that the instrument was evaluated within a relevant educational context.

Content validation was conducted using a Guttman-scale validation sheet, in which each item was rated as either *appropriate* (1) or *not appropriate* (0). The validation assessed four aspects: content relevance, construct appropriateness, language clarity, and the instrument's ability to measure disaster literacy and identify students' conceptual understanding. Quantitative analysis was performed by applying the CVR (*Content Validity Ratio*) and CVI (*Content Validity Index*) procedures proposed by Lawshe (1975). The CVR Score for each test item was determined using the following formula:

$$CVR = \frac{ne - \frac{N}{2}}{\frac{N}{2}}$$

Description:

CVR = Content Validity Ratio

Ne = Number of validators who rated the item as appropriate

N = Total number of validators

CVR scores vary from -1 to 1, with elevated scores signifying greater consensus among experts about the importance of a specific item. In line with Lawshe (1975) guidelines, items that failed to meet the required CVR minimum were updated prior to testing in practice. For this research, there were five evaluators involved, meaning the least acceptable CVR score was 0.99 (Lawshe, 1975). Items that did not meet the minimum CVR criterion were revised based on the validators' recommendations before empirical testing. The overall content validity of the instrument was assessed using the Content Validity Index (CVI), which was calculated by averaging the CVR values across all items. The CVI was computed using the following equation:

$$CVI = \frac{\sum CVR}{k}$$

Where $\sum CVR$ is the sum of the CVR values of the valid items and k is the number of items that met the minimum CVR criterion (Lawshe, 1975). The interpretation criteria for the CVI values used in this study are presented in Table 1.

Table 1. Interpretation Criteria
for CVI Values

CVI Value	Interpretation
0,00-0,33	Invalid
0,34-0,67	Valid
0,68-1,00	Very valid

The interpretation criteria for the CVI values were adapted from Lawshe (1975). The CVI was utilized to assess the general content validity of the tool. An elevated CVI score signifies a stronger consensus among specialists about the significance and representation of the test items with respect to the construct being measured. The instrument was considered to have excellent content validity when the CVI value ranged from 0.68 to 1.00.

RESULT AND DISCUSSION

In accordance with the ADDIE model used in this study, the results presented in this section are outputs from the *Development stage*. In the previous stage, namely *Analyze*, needs analysis was carried out through literature review, student needs questionnaires, and interviews with science teachers to identify the need for the development of disaster literacy instruments. Furthermore, in the *Design stage*, indicators, grids, and 32 multitier diagnostic test instruments were prepared that integrate vibration and wave materials with the context of disasters. The *Development stage* is then focused on validating the content of the instrument through expert assessment using the *Content Validity Ratio (CVR)* and *Content Validity Index (CVI)* approach as the basis for revision before the instrument is tested at the *Implementation stage*. The developed instrument was organized into three disaster literacy aspects, namely knowledge, attitude, and skills. Each aspect

consisted of several sub-aspects representing the indicators of disaster literacy measured by the four-tier diagnostic test. The distribution of the aspects, sub-aspects, and corresponding item numbers is presented in Table 2.

Table 2. Disaster Literacy Aspects and Sub-aspects of the Instrument

Disaster Literacy Aspect	Indicators	Item Numbers
knowledge	Analyze the definition and causes of disasters related to seismic waves.	PG-9, PG-10
	Explain the impacts and hazards of disasters on humans and the environment.	PG-1, PG-2
	Develop disaster mitigation procedures.	PG-4, PG-5
	Develop disaster mitigation and preparedness action plans.	PG-11, PG-12
	Determine appropriate disaster response procedures.	PG-6, PG-7
Attitude	Design rescue and post-disaster medical care procedures.	PG-3, PG-8
	Evaluate the environment and identify potential hazards	PG-13, PG-14
	Synthesize and analyze disaster-related information proactively.	PG-15, PG-16
	Promote the importance of disaster prevention, disaster relief, and evacuation planning.	PG-17, PG-18
	Explain the relationship between disaster prevention and social costs.	PG-19, PG-20
Skills	Organize community-based disaster awareness and prevention activities.	PG-21, PG-22
	Plan evacuation procedures and shelter allocation.	PG-23, PG-24
	Plan evacuation routes and shelter locations during disasters.	PG-25, PG-26
	Plan and participate in disaster prevention and emergency response drills and simulations.	PG-27, PG-28
	Ensure personal safety and assist others in evacuating during disasters.	PG-29, PG-30
	Cooperate during evacuation and shelter management.	PG-31, PG-32

The developed instrument consists of three disaster literacy aspects, namely knowledge, attitude, and skills, which are represented by 18 disaster literacy indicators distributed across 32 multitier diagnostic test items (Table 1). The knowledge aspect comprises eight indicators (12 items), the attitude aspect consists of six indicators (12 items), and the skills aspect includes four indicators (8 items). These indicators were developed to comprehensively assess students' disaster literacy within the context of vibration and wave concepts. After the instrument structure had been established, its content validity was evaluated through expert judgment using the Content Validity Ratio (CVR) and the Content Validity Index (CVI).

Content validity was determined by means of assessment utilizing the Content Validity Ratio (CVR) and the Content Validity Index (CVI). The panel of experts included two higher education instructors and three science educators from junior high schools. Additionally, validation focused on 32 questions that are intended to evaluate disaster literacy in the subjects of vibrations and waves. The evaluation conducted by validators encompasses various factors, including the appropriateness of content, constructs, language, and the instrument's capacity to gauge students' comprehension.

The CVR analysis showed that the values of the developed items ranged from 0.81 to 1.00. According to the criteria proposed by Lawshe (1975), when five validators are involved, the minimum acceptable CVR value is 0.99. Therefore, items with CVR values of 0.99 or higher were considered valid, whereas items below this threshold were revised before proceeding to the next

stage. The items evaluated during the Development stage were derived from the disaster literacy indicators established in the Analysis stage and operationalized into a test blueprint during the Design stage of the ADDIE model. Consequently, the CVR analysis reflects the extent to which each item represents the intended disaster literacy constructs before empirical testing. The summary of the CVR analysis results is presented in Table 3.

Table 3. Recapitulation of Results of Instrument Content Validity

Aspects	Number of Question Items	CVR range	Number of Valid Items	Number of Revised Items
Knowledge	12	0.81-1.00	5	7
Attitude	12	0.89-1.00	7	5
Skills	8	0.90-1.00	6	2
Total	32	-	18	14

The expert judgments showed substantial agreement across all items, as reflected by CVR values ranging from 0.81 to 1.00, indicating a high level of agreement among the validators regarding the content suitability of the instrument. A more detailed examination of the item-level CVR results showed that, within the knowledge aspect, 5 of the 12 items (PG-4, PG-6, PG-7, PG-8, and PG-10) met the minimum CVR criterion. In contrast, the remaining 7 items required revision. In the attitude aspect, 7 of the 12 items (PG-14, PG-15, PG-17, PG-19, PG-20, PG-23, and PG-24) were classified as valid, while the remaining 5 items required revision. For the skills aspect, 4 of the 8 items (PG-28, PG-29, PG-30, and PG-32) satisfied the validity criterion, whereas the remaining 4 items required revision. Overall, 18 of the 32 developed items met the minimum CVR criterion, whereas the remaining 14 items were revised based on the validators' recommendations to improve indicator alignment, wording clarity, disaster context, answer options, and confidence levels in each tier. The knowledge aspect contained the largest number of revised items, suggesting that developing items capable of measuring students' cognitive understanding while simultaneously integrating disaster literacy with vibration and wave concepts was more challenging than developing items for the attitude and skills aspects. However, the relatively large number of revised items does not indicate poor instrument quality; instead, it reflects the effectiveness of the expert validation process in identifying weaknesses and providing constructive feedback to improve the instrument before empirical testing.

The validation results indicate that most of the developed items adequately represent the indicators of disaster literacy across the knowledge, attitude, and action dimension. Although there are still some items that need revision, the results of the validator's assessment show that, in general, the content of the instrument is consistent with the intended construct. Items that do not meet the minimum score requirements are not discarded directly, but are corrected based on the validator's suggestions, especially in terms of phrasing clarity, the suitability of indicators, the accuracy of the disaster context, as well as the preparation of answer choices and confidence levels in each tier. The revision process is needed so that each item can measure disaster literacy and identify student understanding more accurately.

The findings of this study are consistent with Almanasreh *et al.* (2018), who stated that content validity is a fundamental stage in instrument development because it ensures that each item appropriately represents the intended construct based on expert judgment. Instruments that have undergone expert validation and subsequent revision generally provide more accurate and representative measurements of the target construct. Although all assessment instruments require content validation before use, the validation process is particularly critical for four-tier diagnostic tests because each item consists of four interrelated components: the answer, the reasoning, the confidence in the answer, and the confidence in the reasoning (R. E. Putri & Subekti, 2021). Any weakness in one of these components may reduce the instrument's ability to diagnose

students' conceptual understanding and misconceptions accurately. Therefore, each tier should be carefully evaluated during the content validation process.

The CVR and CVI values presented in this study were calculated based on the validators' initial assessments of the instrument. Along with assigning relevance ratings, the validators also provided qualitative comments and suggestions for improvement. These comments were subsequently used to revise items that did not meet the minimum CVR criterion as well as items requiring refinement in terms of wording, indicator alignment, disaster context, answer options, and confidence levels. The revised version of the instrument was then prepared for the subsequent Implementation stage of the ADDIE model, where empirical testing was conducted. The suggestions and revisions resulting from the expert review are presented in Tables 4 and 5.

Table 4. Suggestions and Revisions from Validator 1

Field of Expertise
Natural Science Teacher
Suggestion
The editorial of the stimulus questions is clarified and simplified to make it easier for junior high school students' to understand.
Revision
The wording of the question stimulus was revised to improve clarity, readability, and the logical flow of information. The revisions included: (1) simplifying the opening sentence by changing "Observe the figure above showing how an earthquake occurs!" to "Observe the figure above showing how earthquakes occur!" to eliminate unnecessary repetition; (2) revising the sentence "Earthquake energy is released at the focus, and seismic waves move toward the epicenter" to "Earthquake energy is released at the focus, after which seismic waves propagate toward the epicenter" to present the sequence of events more clearly; and (3) replacing the expressions "people" with "the community" and "make buildings collapse" with "cause buildings to collapse" to achieve a more formal style and improve comprehensibility for junior high school students'.

As shown in Table 4, Validator 1 mainly recommended improving the wording and readability of the question stimuli. The revisions focused on simplifying sentence structures, improving the logical flow of information, and using more appropriate terminology for junior high school students'. These revisions were intended to improve students' understanding of the questions without changing the intended disaster literacy constructs being measured.

Table 5. Suggestions and Revisions from Validator 2

Field of Expertise
Physics Education Lecturer
Suggestion
General Suggestions
1) The items in the knowledge aspect should be arranged from lower-order to higher-order cognitive levels.
2) Disaster mitigation should be organized into three phases before, during, and after a disaster to provide a more systematic conceptual framework for the questions.
Specific Suggestions
1) The answer choices for Item PG-12 should be made more detailed.
2) The wording of the stimulus in Item PG-18 should be expanded to provide clearer context.
3) The Tier 1 answer choices in Item PG-11 are insufficient and should be revised.
4) Item PG-19 is not aligned with the disaster literacy indicator and should be revised accordingly.

-
- 5) In Item PG-16, the Tier 3 answer choices B and C are identical and should be differentiated.
 - 6) Item PG-27 should be revised because the evacuation simulation scenario is unrealistic, as a simulation cannot reasonably be conducted when a tsunami is expected to occur within 15 minutes. The current item is therefore ambiguous.
 - 7) In Item PG-31, the question should use either the term "action" or "skill", rather than combining both in the same statement.
-

Revision

General Suggestions

- 1) The sequence of items in the knowledge aspect was reorganized according to Bloom's Taxonomy, progressing from lower-order to higher-order cognitive levels. This revision provides a more systematic and gradual sequence for students' when completing the instrument.
- 2) The disaster mitigation items were revised by categorizing the mitigation contexts into the phases before, during, and after a disaster. Specifically, Item PG-15 was revised to represent the pre-disaster phase, PG-7 the during-disaster phase, and PG-11 the post-disaster phase. These revisions resulted in a more systematic presentation of disaster mitigation content in accordance with the validator's recommendations.

Specific Suggestions

- 1) The answer choices in Item PG-12 were revised to provide more detailed and sequential emergency response procedures. Specifically, the Tier 1 options were expanded to describe complete first-aid actions rather than isolated responses, allowing each alternative to represent a distinct procedure. In addition, the Tier 3 reasoning option was refined by incorporating the concept of emergency triage, emphasizing that victims with severe injuries should be prioritized in situations with a high risk of damage. These revisions improved the clarity, specificity, and alignment of the item with the intended disaster literacy indicator.
 - 2) Item PG-18 was revised by adding contextual information to the stimulus to improve clarity. Evidence of the revision: The following statements were added: *"In coastal areas, the risk of tsunamis is relatively high; therefore, communities need to understand early warning signs and appropriate self-evacuation procedures,"* and *"Based on this situation, students' are expected to develop awareness and support tsunami mitigation education within their communities."*
 - 3) The Tier 1 answer choices for PG-11 were revised to improve clarity and specificity. Evidence of the revision: Option B was revised to *"Use the radio to alternately communicate essential information and report the victims' locations and conditions to facilitate rescue operations."* Option C was revised to *"Wait until the cellular network is restored before coordinating rescue efforts."* Option D was revised to *"Use the radio only to receive information without reporting the victims' conditions in the field."*
 - 4) Item PG-19 was revised to better align with the disaster literacy indicator. Evidence of the revision: The stimulus was expanded by adding the statement *"Earthquakes can cause structural damage, casualties, and social and economic losses. Disaster mitigation aims to reduce these impacts."* The question was also revised to *"Which attitude demonstrates an understanding of the relationship between pre-earthquake mitigation efforts and the reduction of social and economic losses?"*
 - 5) The Tier 3 reasoning options for PG-16 were revised to eliminate duplicate alternatives. Evidence of the revision: Option B was changed to *"Disaster preparedness is influenced more by personal experience than by scientific information,"* while Option C was revised to *"Scientific information, such as vibration frequency, can help estimate hazard levels and is therefore important to understand."*
-

-
- 6) Item PG-27 was revised to remove ambiguity in the evacuation simulation context. Evidence of the revision: The statement *“The school schedules an evacuation simulation”* was revised to *“The school schedules an evacuation simulation to prepare students’ for real emergency situations.”* In addition, Option A was revised to *“Conduct the evacuation simulation with a target evacuation time of less than 15 minutes using the designated evacuation route to match the estimated tsunami arrival time.”*
 - 7) Item PG-31 was revised by simplifying the wording of the question to avoid using the terms *“action”* and *“skill”* simultaneously. Evidence of the revision: The phrase *“The appropriate skill-related action is ...”* was changed to *“The most appropriate action to support evacuation coordination is ...”* The stimulus was also supplemented with the statement *“Each evacuation team is equipped with a two-way radio for communication and coordination.”*
-

Based on the suggestions and revisions presented in Tables 4 and 5, the validators' feedback primarily focused on improving the content quality of the instrument, particularly the clarity of the question wording, the alignment of the items with the disaster literacy indicators, the appropriateness of the cognitive levels based on Bloom's Taxonomy, and the construction of answer choices and reasoning across the four tiers. The validators also recommended using more realistic disaster contexts and organizing mitigation-related items according to the phases before, during, and after a disaster to ensure more systematic coverage of the indicators.

These recommendations indicate that the validators evaluated not only the overall content validity of the instrument but also its pedagogical and contextual quality. Clear wording is essential to ensure that students correctly understand the intended meaning of each question. At the same time, the alignment between indicators and cognitive levels helps ensure that every item accurately measures the intended competencies. Furthermore, incorporating realistic disaster contexts strengthens the relationship between science concepts and real-life situations that students' may encounter, particularly in earthquake- and tsunami-prone areas.

The findings support the opinion of Ramadan *et al.* (2026). Expert validation not only functions to determine the feasibility level of an instrument, but also expert comments are an important part of the content validity process because they are able to increase the clarity, relevance, and representation of each item to the measured construct. Based on these inputs, revisions are made to items that have not met the minimum CVR value as well as to items that have received notes from validators. The revision includes improving the editorial, adjusting indicators, improving the choice of answers and reasons in each tier, as well as improving the disaster context so that the instrument becomes more representative in measuring disaster literacy and identifying students' understanding. Input from validators is the basis for the implementation of formative evaluation at the *Development* stage in the ADDIE model. This evaluation is carried out by revising items that have not met the minimum CVR value and improving items that still have weaknesses based on validator comments. Thus, the resulting instrument has better content suitability before proceeding to the *Implementation* stage to obtain evidence of empirical validity. Within the ADDIE framework, the expert feedback constituted the formative evaluation conducted during the Development stage. The quantitative evidence obtained from the CVR and CVI analyses, together with the qualitative comments from the validators, guided the revision of the instrument before it advanced to the Implementation stage. This process ensured that the instrument had satisfactory content validity before empirical testing.

Content Validity Index (CVI) Analysis Results

After the CVR analysis, the Content Validity Index (CVI) was calculated to evaluate the overall content validity of the instrument. Before calculating the CVI, the expert ratings were summarized based on the number of appropriate (1) and not appropriate (0) responses for each

disaster literacy aspect, as presented in Table 6. These ratings formed the basis for the calculation of the Σ CVR and CVI values presented in Figure 1.

Table 6. Summary of Expert Ratings for Each Disaster Literacy Aspect

Disaster Literacy Aspect	Number of Items	Yes (1)	No (0)	Σ CVR	CVI
Knowledge	12	1295	25	11.615	1
Attitude	12	1310	10	11.81	1
Skills	8	863	17	7.68	0.96

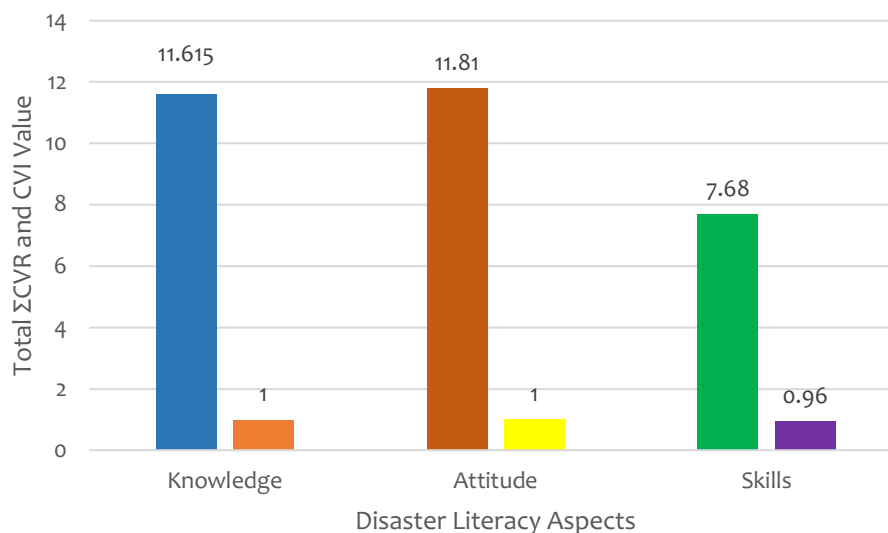


Figure 1. The Σ CVR and CVI values on each aspect of the instrument

Based on Table 6 and Figure 1, the content validity analysis showed a high level of agreement among the five expert validators across all disaster literacy aspects. The knowledge aspect, consisting of 12 items, received 1,295 "Yes (1)" responses and 25 "No (0)" responses, resulting in a Σ CVR value of 11.615 and a CVI of 1.00. The attitude aspect also consisted of 12 items and obtained 1,310 "Yes (1)" responses and 10 "No (0)" responses, producing the highest Σ CVR value of 11.81 with a CVI of 1.00. Meanwhile, the skills aspect, comprising 8 items, received 863 "Yes (1)" responses and 17 "No (0)" responses, resulting in a Σ CVR value of 7.68 and a CVI of 0.96. Overall, the instrument achieved a CVI value of 0.972. According to the interpretation criteria adapted from Lawshe (1975), the overall CVI of 0.972 falls within the very valid category, indicating that the developed instrument possesses excellent content validity. The high level of agreement among the expert validators demonstrates that the instrument adequately represents the intended disaster literacy construct across the knowledge, attitude, and skills dimensions. Although the skills aspect obtained a slightly lower CVI than the other aspects, its value (0.96) still exceeded the recommended minimum criterion, indicating that all aspects of the instrument were considered content valid.

The consistently high CVI values across the three disaster literacy aspects indicate that the revisions made based on the validators' suggestions successfully improved the quality and relevance of the instrument. These findings confirm that the objectives of the Development stage in the ADDIE model were achieved by producing an instrument with satisfactory content validity. Therefore, the revised instrument was considered appropriate to proceed to the Implementation stage, where empirical data from junior high school students were collected for further validity and reliability analyses.

CONCLUSION

The expert validation results showed that the CVR values ranged from 0.81 to 1.00. Of the 32 developed items, 18 met the minimum CVR criterion, whereas the remaining 14 items were revised based on the validators' recommendations regarding indicator alignment, wording clarity, disaster context, and the construction of answer choices and confidence levels across the four tiers. The overall CVI value of 0.972 indicates that the instrument possesses excellent content validity. This study contributes a validated four-tier diagnostic test that integrates disaster literacy with vibration and wave concepts, providing a comprehensive instrument for assessing students' conceptual understanding and misconceptions. One of the main challenges encountered during the development process was ensuring consistency among the four interconnected tiers while maintaining appropriate disaster literacy indicators and realistic disaster contexts. After the revision process, the instrument was considered suitable for empirical testing with junior high school students. Future studies are recommended to examine its empirical validity, reliability, and practicality using larger and more diverse student samples.

DECLARATION OF GENERATIVE AI

During the preparation of this work, the author used ChatGPT to assist with paraphrasing sentences, grammar checking, and improving the clarity of the manuscript. The author also used Grammarly to assist with grammar and language editing. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

CONFLICTING INTEREST STATEMENT

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING INFORMATION STATEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHORSHIP CONTRIBUTION STATEMENT

Apriyani: Methodology, Data Curation, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Henny Johan, Rahmawati, Muchammad Farid, Sutarno, Aprina Defianti:** Supervision and Guidance.

REFERENCES

- Almanasreh, E., Moles, R., & Chen, T. F. (2018). Research in social and administrative pharmacy evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 1–8. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Amelia, R. N., Hasmawati, & Aswan, M. (2023). Model learning by doing terhadap kemampuan literasi bencana konteks gempabumi di mts negeri 1 pulau morotai. *Jurnal Spatial Wahana Komunikasi dan Informasi Geografi*, 23(2), 161–169. <https://doi.org/10.21009/spatial.232.08>

- Anisah, N. (2019). Model sekolah aman bencana dalam upaya mewujudkan pendidikan karakter di MIN 1 Bantul. *Literasi: Jurnal Ilmu Pendidikan*, X(1), 9–20. [https://doi.org/10.21927/literasi.2019.10\(1\)](https://doi.org/10.21927/literasi.2019.10(1))
- Berlianawati, A. W., Prasetya, S. P., Marzuqi, M. I., & Prasetyo, K. (2025). Efektivitas dan kelayakan LKPD kebencanaan untuk meningkatkan literasi kebencanaan siswa SMPN 4 Sidoarjo. *Jurnal Dialektika Pendidikan IPS*, 5(1), 285–292. <https://ejournal.unesa.ac.id/index.php/PENIPS/index>
- Branch, R. M. (2009). *Instructional design-the ADDIE approach*. New York: Springer.
- Febriana, H., & Margereta, S. (2026). Pengembangan instrumen four tier diagnostic test untuk mengidentifikasi profil miskonsepsi mahasiswa pada materi biologi sel. *Jurnal edu research Indonesian Institute for Corporate Learning and Studies (IICLS)*, 7(1), 3802–3812. <https://doi.org/10.47827/jer.v7i1.2514>
- Fiskawarni, T. H., Rahmawati, Widiasih, & Saad, R. (2024). The design and validation of the four tier test instrument for energy literacy using the rasch model analysis. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, 9(1), 74–87. <https://doi.org/10.26737/jipf.v9i1.4717>
- Halik, M. F. Al, & Septiana, L. (2022). Analisa data untuk prediksi daerah rawan bencana alam di jawa barat menggunakan algoritma k-means clustering. *Journal of Information System, Applied, Management, Accounting and Research*, 6(4), 856–870. <https://doi.org/10.52362/jisamar.v6i4.939>
- Handayani, E. P., Kustati, M., Amelia, R., & Gusmirawati. (2024). Program pelatihan satuan pendidikan aman bencana (spab). *Jurnal Pengabdian Kepada Masyarakat*, 2(1), 82–92. <https://journal.staittd.ac.id/index.php/inv/article/view/177>
- Hilft, B. E. (2023). *World Risk Report 2023*. Berlin: Bündnis Entwicklung Hilft.
- Islami, M., Utaya, S., & Purwanto. (2025). Model guided inquiry learning terhadap kemampuan berpikir spasial siswa sebagai penguat literasi kebencanaan. *Journal of Education Action Research*, 9(1), 87–95. <https://doi.org/10.23887/jea.v9i1.90623>
- Johan, H., Mayub, A., & Wardana, R. W. (2021). Integrasi pendidikan mitigasi dalam konten pembelajaran sains sebagai upaya membangun kesadaran mitigasi dari jenjang sekolah. *Jurnal Ilmiah Pengembangan Dan Penerapan IPTEKS*, 19(01), 186–196. <https://doi.org/10.33369/dr.v19i1.13790>
- Junianto, D., & Hendriani, D. (2024). Pengaruh model simulasi bencana terhadap nilai sikap kesiapsiagaan siswa pada mata pelajaran ilmu pengetahuan sosial. *CENDEKIA: Jurnal Penelitian Dan Pengkajian Ilmiah*, 1(1), 45–52. <https://doi.org/10.62335/kozhsw43>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Psychology*, 28(1), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Nainggolan, Y. A., Wahyudi, R., & Januar. (2026). Teori validasi model dan instrumen penelitian pengembangan. *Jurnal Edu Research Indonesian Institute For Corporate Learning And Studies (IICLS)* Page, 7(2), 2682–2695. <https://doi.org/10.47827/jer.v7i2.2870>
- Polit, D. F., & Beck, C. T. (2006). The content validity index : are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, 29, 489–497. <https://doi.org/10.1002/nur>

- Priyasmika, R., & Sholichah, N. (2022). Analisis miskonsepsi pada materi hidrolisis garam menggunakan instrumen tes diagnostik four tier. *Jurnal Riset Teknologi Dan Inovasi Pendidikan (Jartika)*, 5(1), 19–28. <https://journal.rekarta.co.id/index.php/jartika>
- Putri, D. T., Umah, A. M., Fadilah, A. A., Cahya, E. D., Ristanti, A. N., & Susilo, A. (2025). Dampak literasi kebencanaan terhadap kesiapsiagaan peserta didik dalam mitigasi bencana banjir: Studi di MTSN 4 Magetan. *Jurnal Pendidikan Geografi Undiksha*, 13(1), 78–86. <https://doi.org/10.23887/jjpg.v13i1.81752>
- Putri, R. E., & Subekti, H. (2021). Analisis miskonsepsi menggunakan metode four-tier certainty of response index: studi eksplorasi di SMP Negeri 60 Surabaya. *Pensa E-Jurnal: Pendidikan Sains*, 9(2), 220–226. <https://doi.org/10.26740/pensa.v9i2.38279>
- Rahmawati, Widiasih, Marisda, D. H., & Riskawati. (2023). Using four-tier test to identify prospective elementary teacher students' misconception on electricity topic. *Jurnal Penelitian Pendidikan IPA*, 9(10), 7793–7802. <https://doi.org/10.29303/jppipa.v9i10.3272>
- Ramadhan, H., Basri, I. Y., Maksum, H., & Rifdarmon. (2026). Pengembangan e-book pada mata kuliah teknologi sensor dan kontrol elektronik otomotif. *Tsaqofah: Jurnal Penelitian Guru Indonesia*, 6(4), 3434–3448. <https://doi.org/https://doi.org/10.58578/tsaqofah.v6i4.10553>
- Safitra, L., Susiyanto, Rifa'i, Yuliani, H., & Ayu, E. T. (2023). Penguatan modal sosial keluarga dalam kegiatan mitigasi bencana di kelurahan Malabero kecamatan Teluk Segara Kota Bengkulu. *JAMS: Jurnal Abdimas Serawai*, 3(3), 139–151. <https://doi.org/10.36085/jams.v3i3.5897>
- Sudrajat, & Wibowo, S. (2016). Pemahaman tentang manajemen bencana alam siswa sekolah menengah pertama. *JIPSINDO (Jurnal Pendidikan Ilmu Pengetahuan Sosial Indonesia)*, 3(2), 168–189. <https://doi.org/10.21831/jipsindo.v3i2.11697>
- Susmita, A. R., Darmawi, E., & Suri, E. W. (2023). Analisis pendekatan mitigasi bencana alam di badan nasional penanggulangan bencana kota Bengkulu. *Jurnal Ilmiah Administrasi Publik*, 1(2), 9–18. <https://dx.doi.org/10.32663/md.v1i2.3812>
- Tahmidaten, L., & Krismanto, W. (2019). Implementasi pendidikan kebencanaan di Indonesia (sebuah studi pustaka tentang problematika dan solusinya). *Lectura: Jurnal Pendidikan*, 10(2), 136–154. <https://doi.org/10.31849/lectura.v10i2.3093>
- Uswatun, & Mubarak, S. (2024). Pengembangan instrumen tes diagnostik four-tier untuk mengidentifikasi miskonsepsi materi konsep mol dan stoikiometri. *Al Kawnu: Science And Local Wisdom Journal*, 4(1), 34–49. <https://doi.org/10.18592/ak.v4i1.13838>