https://doi.org/10.30599/7vqrzc10

Content Validity of Augmented Reality Flashcards on Fluid Dynamics Material Using Lawshe's Content Validity Ratio and an Integrated Approach

Mutiara Putri[1], Hamdi Akhsan[2*], Sardianto Markos Siahaan[3], Leni Marlina[4]
[1,2,3,4] Universitas Sriwijaya
[*] E-mail: *hamdiakhsan@fkip.unsri.ac.id*

Abstract

This study evaluates the content validity of Augmented Reality (AR) flashcards for teaching Fluid Dynamics using a mixed-methods approach. Ten physics experts assessed 50 items across five dimensions. Validity was measured using Lawshe's Content Validity Ratio (CVR; threshold ≥0.80), Gwet's AC1 for inter-rater agreement, and bootstrap sensitivity analysis. All items met statistical validity criteria: 92% achieved perfect agreement (CVR = 1.00) and 8% met the threshold (CVR = 0.80). The overall mean CVR was 0.98 (95% CI: 0.97–1.00), with Gwet's AC1 = 0.98, indicating excellent expert consensus. Sensitivity analysis confirmed result stability despite the limited panel. Qualitative feedback guided iterative improvements to visual design and equation annotations. The instrument demonstrates robust, methodologically sound content validity and is highly suitable for evaluating immersive physics learning tools. The combined CVR, AC1, and sensitivity framework is recommended as a standard for validating technology-enhanced educational media.

Keywords: Content Validity Ratio (Cvr); Augmented Reality Flashcards; Fluid Dynamics; Instrument Validation; Physics Education; Educational Technology

INTRODUCTION

Teaching physics on the topic of fluid dynamics still faces significant challenges regarding students' conceptual understanding. Studies indicate that misconceptions regarding fluid dynamics remain prevalent due to a lack of learning materials that interactively visualize abstract processes (Tiara Ganish et al., 2025). Students' limited ability to use multiple representations when explaining fluid dynamics poses a major obstacle to developing comprehensive critical thinking skills (Erlangga et al., 2024). Digital flashcards have emerged as a solution for visualizing these concepts; however, their effectiveness in stimulating reasoning ability is highly dependent on the precision of message design and the accuracy of the content presented (Yusuf et al., 2021). Yet, the validity of the development instruments is often not statistically tested with a strong foundation (Sarnita et al., 2025). Based on a review of articles on physics media development in Indonesia, it was found that the majority of studies still rely on descriptive validity (e.g., percentage of agreement or average Likert scores) without converting them to standard statistical critical values (Ridwan et al., 2021).

Content validity is a critical foundation in the development of assessment instruments and instructional materials. The Content Validity Ratio (CVR) method, developed by Lawshe (1975), is recommended for quantitatively measuring the degree of expert agreement, particularly when the number of panelists is limited (Ayre & Scally, 2014). CVR values range from -1 to +1, and the decision is determined by comparing the calculated results to Lawshe's critical value table, adjusted for the number of panelists. The advantage of the CVR over descriptive approaches is its ability to provide an objective threshold: an item is deemed valid only if expert agreement exceeds the probability threshold for chance agreement (Nabilah et al., 2025). In this context, the use of CVR provides justification that expert agreement does not occur due to chance but rather stems from the intrinsic quality of the developed media (Romero Jeldres et al., 2023). The difficulty in quantifying this threshold poses a major challenge in ensuring message accuracy in immersive technology-based media such as AR, which involves high cognitive and visual complexity.

The five validation dimensions used in this study (1) Media, (2) Content, (3) Critical Thinking Skills (CTS), (4) Language, and (5) Design were systematically selected to ensure construct alignment between AR technology features and the pedagogical objectives of physics. The Media and Design dimensions are based on Cognitive Load Theory, which emphasizes the reduction of extrinsic cognitive load through visual hierarchy, interface simplification, and AR navigation ergonomics so that students' working memory is not burdened by graphical distractions. The Content and Curriculum Dimensions directly align with the national curriculum and operationalize indicators of higher-order critical thinking (analysis, evaluation, and synthesis of fluid concepts), ensuring that content validity measures not only the accuracy of the content but also the media's capacity to trigger deep cognitive processing. The Language Dimension ensures precision in scientific terminology and unambiguous sentence structure, serving to prevent distortions in understanding caused by linguistic ambiguity that often leads to misconceptions in physics. The integration of these five dimensions transforms the validation process from a mere checklist review into an inferential evaluation that tests the alignment between theoretical constructs, visual representations, and learning objectives.

Although the urgency of statistical validation has been widely discussed, the application of CVR in the development of physics learning materials in Indonesia remains limited. Many researchers tend to use Likert scales without referring to clear critical values, resulting in ambiguous validity that is difficult to defend in scientific forums (Zavira & Sari, 2023). This situation is critical because such weaknesses risk allowing misconceptions to persist if the materials are used directly without rigorous validation screening through statistical thresholds (Afifah Aulia Zayrin et al., 2025). Therefore, by applying the CVR method, researchers can make transparent decisions regarding whether items should be retained, revised, or discarded based on quantitative evidence (Wahyu Prananto et al., 2022).

This study aims to optimize the validation procedure through rigorous mathematical verification, to ensure that every item in the AR flashcard media is truly essential and free from panelists' subjective bias. Through this approach, it is hoped that the limitations of applying the CVR method in the development of physics educational media in Indonesia can be addressed. Based on this gap, this study aims to fill it by conducting rigorous mathematical verification of the content validity of the flashcard media, thereby producing a learning instrument with scientific legitimacy and high reliability (Bhakti et al., 2023). With this approach, it is hoped that strong empirical evidence regarding the suitability of the media will be obtained before it is implemented in actual learning.

However, with the ongoing integration of AR in physics education, there remains a significant research gap in the rigors statistical validation of this media, particularly in the Indonesian context. Most existing work focuses heavily on descriptive validity (e.g., average Likert scores or percent agreement), which lacks the statistical power to rule out chance agreement. Moreover, there are few studies that adopt a combination of multiple robust statistical metrics to validate immersive media.

However, with the ongoing integration of AR in physics education, there is a big research gap in the rigorous (sebelumnya: rigors) statistical validation of this media, especially in the Indonesian context. Most of the existing work has focused heavily on descriptive validity (e.g., average Likert scores or percent agreement), which lacks the statistical power to rule out chance agreement. Also, few studies use a combination of multiple robust statistical metrics to validate immersive media. The novelty of the current study is the integrated validation framework. This study is the first to combine Lawshe's Content Validity Ratio (CVR) to ascertain essentiality of items, Gwet's AC1 to avoid prevalence bias in inter-rater agreement and bootstrap sensitivity analysis to verify the stability of results despite small expert panel. This multi-method statistical approach provides a more defensible, objective, and replicable standard for validating technology-enhanced educational media, ensuring that the AR flashcards are not only visually engaging but also pedagogically and cognitively sound.

METHOD

This study is an instrument validation study that forms part of a research and development project. The primary focus of this study is to test the content validity of digital flashcards on fluid dynamics material prior to conducting a pilot test with students. A quantitative approach was used to objectively measure the degree of expert agreement using Content Validity Ratio (CVR) statistical analysis (Romero Jeldres et al., 2023). The validators in this study consisted of ten experts (expert judgment) selected using purposive sampling based on criteria of competence in the fields of physics education and media development. The expert panel comprised 5 physics education lecturers and 5 high school physics teachers. The selection of exactly 10 panelists was methodologically deliberate to meet Lawshe's critical value threshold ($CVR \geq 0.80$ at $\alpha = 0.05$) while balancing theoretical and practical expertise. This combination of academic and practitioner validators was chosen to ensure the ecological validity of the media, meaning the media is not only content-accurate but also ready for implementation (Amelia & Nasution, 2025). The selection of 10 panelists (5 academics and 5 practitioners) was based on their specific expertise in the field of physics education integrated with Augmented Reality (AR) technology, which is limited in the local context. This combination of theoretical experts (faculty members) and implementation experts (teachers) ensures broad content coverage and the ecological validity of the media, making it ready for implementation in real classrooms.

Research data were collected using a validation sheet systematically designed based on adapted educational media evaluation standards to capture the panelists' professional assessments.

https://doi.org/10.30599/7vqrzc10

The instrument is organized into five main aspects: (1) Media Suitability, (2) Content Suitability, (3) Critical Thinking Skills, (4) Language, and (5) Media Design. Each statement item uses a dichotomous scale with two response options: “Essential” (Yes) and “Not Essential” (No). This specific scale was chosen to facilitate the calculation of the CVR coefficient, which strictly requires categorical data. In addition to the quantitative assessment, the validation sheet also provided a column for qualitative suggestions and feedback, which served as the basis for formative revisions to the media components.

Data Analysis Techniques

The Content Validity Ratio (CVR) analysis in this study followed Lawshe’s (1975) procedure to quantitatively measure the degree of expert agreement on dichotomous data. The CVR value was calculated using the formula:

$$CVR=ne\text{-}N2N2$$

Notes:

- CVR= Content Validity Ratio for each item
- ne= Number of panelists who rated the item as “Essential”
- N= Total number of validator panelists (N=10)

The expert rating data were analyzed using the Content Validity Ratio (CVR) formula developed by Lawshe (1975). The CVR formula is used to measure the proportion of validators who rated an item as “essential” compared to the probability of agreement by chance (Slater P et al., 2025). CVR values range from -1.00 to +1.00. The higher the CVR value, the stronger the experts’ agreement that the item is relevant and representative of the construct being measured. To provide a more comprehensive picture of inter-rater agreement, Gwet’s AC1 coefficient was also calculated as a more stable alternative to Kappa regarding prevalence bias.

Table 1. Item Validity Decision Criteria

Category	CVR Criteria	Action
Valid	CVR ≥ 0.80	Item retained without substantial revision
Marginal	0,00 ≤ CVR < 0.80	Item requires revision based on qualitative feedback from panelists
Invalid	CVR < 0.00	Item removed from the instrument

(Romero Jeldres et al., 2023)

RESULTS AND DISCUSSION

Result

1. **Construction of the Validation Instrument**

Content validity is a fundamental stage in instrument development to ensure that items represent the construct being measured in a relevant and comprehensive manner (Roy et al., 2023). In this study, the content validity of Augmented Reality (AR)-based flashcards on Fluid Dynamics was tested using Lawshe’s (1975) Content Validity Ratio (CVR) method, involving ten expert panelists. This approach was chosen due to its ability to quantitatively measure inter-expert agreement on dichotomous data, thereby minimizing the subjective bias often found in conventional descriptive approaches. Based on the critical value table for the number of panelists (N) = 10 at a significance level of α = 0.05, the minimum CVR threshold is 0.80. An item is considered statistically valid if the CVR value is ≥ 0.80.

Table 2. Validation Matrix for Augmented Reality-Based Flashcard Media

Aspect	Indicator	Item
Media (M)	Media Display	M1, M2, M3, M4, M5, M6, M7
	Media in Education	M8, M9
	Student Engagement with Media	M10, M11, M12
Material (MT)	Curriculum	MT1, MT2, MT3
	Learning	MT4, MT5, MT6
	Content	MT7, MT8
	Interaction	MT9
Critical Thinking Skills (CTS)	Providing Simple Explanations	CTS1, CTS2, CTS3
	Building Basic Skills	CTS4, CTS5
	Drawing Conclusions	CTS6, CTS7, CTS8
	Providing Further Explanations	CTS9, CTS10

	Developing Strategies and Tactics	CTS11, CTS12
Language (L)	Readability	L1, L2, L3
	Clarity of Information	L4
	Compliance with Indonesian Language Rules	L5, L6
	Effective and Efficient Use of Language	L7, L8, L9
Design (D)	Presentation Order	D1
	Cover and Flashcard Content Design	D2, D3, D4
	Font Usage	D5, D6
	Layout	D7, D8

The validation instrument in Table 2 maps 50 items into five assessment dimensions Media, Content, Critical Thinking Skills, Language, and Design based on the principle of construct mapping, which ensures that each indicator is clearly derived from the theoretical construct and operationalized in a measurable manner. The proportional distribution of items (9-12 items per aspect) reflects a multidimensional validation approach to ensure comprehensive evaluation coverage without sacrificing the depth of assessment in each dimension (Slater & Hasson, 2025).

The application of the construct mapping principle in the development of this instrument ensures that each item has a strong theoretical foundation, so that the validation results are not merely descriptive but also provide inferential insights into the overall quality of the media. The relationship between theoretical constructs and operational indicators in each dimension allows researchers to specifically identify areas requiring revision, both in technical and pedagogical aspects. This aligns with contemporary validation approaches that emphasize the importance of alignment between measurement objectives and the constructs being assessed to produce instruments with adequate content and construct validity (Nisrina et al., 2022; Setiawan et al., 2024). Thus, this instrument matrix serves not merely as a checklist, but as an analytical framework that guides a systematic and focused evaluation process.

In the Media and Design dimension, the inclusion of 9-12 items per aspect allows for an in-depth assessment of the visual aspects, interactivity, and cognitive ergonomics of AR media. The visualization of the abstract and dynamic concept of Fluid Dynamics requires accurate graphical representation and intuitive navigation to avoid overloading students' working memory. Research by (Arifta Nurjanah et al, 2024) indicates that suboptimal AR interface design can increase extraneous cognitive load, which in turn reduces students' capacity to process essential information. Therefore, indicators such as layout, font usage, and presentation sequence in this instrument play a strategic role in ensuring that visual aesthetics remain subordinate to instructional objectives, so that the media is not only visually appealing but also effective in supporting knowledge construction (Markos Siahaan et al., 2023).

The dimensions of content and critical thinking skills (CTS) form the core of pedagogical validation, given that the primary purpose of this media is to facilitate conceptual understanding while fostering the development of higher-order cognitive competencies. Assessment items covering curriculum alignment, content depth, and CTS indicators such as drawing conclusions and providing further explanations are designed to measure the extent to which the media can stimulate meaningful cognitive activities. Findings (Dewi, P.S., & Kuswanto, 2023) confirm that interactive science media that integrate critical thinking indicators into their design tend to be more effective in enhancing students' analytical and evaluative skills regarding physical phenomena. Additionally, linguistic aspects including readability, clarity of information, and adherence to language conventions form the foundation of accurate scientific communication; linguistic ambiguity in explanations of fluid concepts risks reinforcing misconceptions if not rigorously validated from the early development stage (López-Belmonte et al., 2022).

Holistically, the multidimensional approach in this instrument reflects the modern educational media evaluation paradigm, which no longer treats technical aspects, content, and cognitive impact as separate entities. The integration of these five assessment dimensions enables more robust triangulation of validation data, allowing decisions regarding the media's suitability to be made based on comprehensive empirical evidence. (Dewi, P.S., & Kuswanto, 2023) assert that a holistic validation instrument structured based on construct mapping possesses higher predictive power regarding the effectiveness of media implementation in real classroom contexts. Thus, the instrument matrix in Table 2 not only meets the methodological standards of development research but also contributes to strengthening the theoretical framework for evaluating immersive technology-based learning media in science education.

2. Content Validity Analysis Based on the Content Validity Ratio

Based on the evaluations of 10 experts, the Content Validity Ratio (CVR) was calculated for each item. The distribution of CVR values per item and a summary by aspect are presented in Tables 3 and 4 below.

Table 3. Results of the CVR Analysis of the Flashcard Media Validation Instrument

CVR	Item Code	Total
0.80	M3, M7, MT7, CTS8	4
1.00	M1, M2, M4, M5, M6, M8, M9, M10, M11, M12, MT1, MT2, MT3, MT4, MT5, MT6, MT8, MT9, CTS1, CTS2, CTS3, CTS4, CTS5, CTS6, CTS7, CTS9, CTS10, CTS11, CTS12, L1, L2, L3, L4, L5, L6, L7, L8, L9, D1, D2, D3, D4, D5, D6, D7, D8	46

Table 4. Summary of CVR Results by Validation Aspect

Validation Aspect	Numbers of Items	CVR Range	Average CVR	Valid Items (CVR ≥ 0.80)	Validity Percentage
Media	12	0.80 – 1.00	0.96	12	100%
Content	9	0.80 – 1.00	0.97	9	100%
Critical Thinking Skills	12	0.80 – 1.00	0.98	12	100%
Language	9	1.00 – 1.00	1.00	9	100%
Design	8	1.00 – 1.00	1.00	8	100%
Total	50	0.80 – 1.00	0.98	50	100%
95% Confidence Interval (CI)	-	-	0.97 - 1.00	-	-
Gwet's AC1			0.98		Very good

Based on Table 4, the overall average CVR was 0.98 (95% CI: 0.97–1.00) with Gwet's AC1 of 0.98, indicating a very high level of content validity and inter-rater agreement in the "Very Good" category. All 50 items exceeded the critical CVR threshold of ≥ 0.80, with 46 items (92%) achieving full agreement (CVR = 1.00). These findings confirm that the instrument is aligned with the content blueprint and free from panelist subjectivity bias.

https://doi.org/10.30599/7vqrzc10

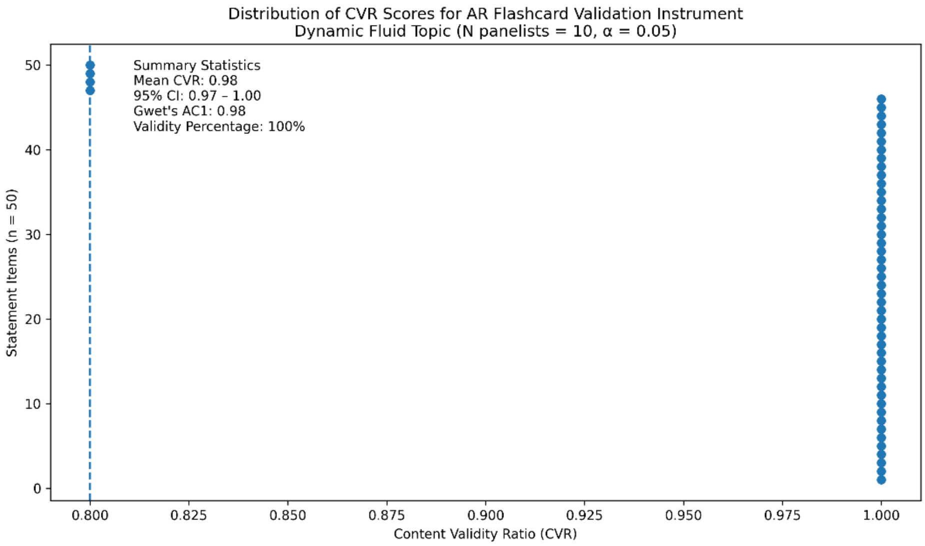


Figure 1. Distribution of CVR Values for the AR Flashcard Validation Instrument

As shown in Figure 1, the distribution of CVR values indicates a clear polarization between the two groups of items. A total of 46 items (92%) received a CVR value of 1.00, reflecting unanimous agreement among all panelists, while 4 items (8%) fell exactly at the critical threshold of CVR = 0.80, marked by the dashed line. No item fell below this critical threshold, so all 50 items were deemed valid. Descriptive statistics show an average CVR of 0.98 with a 95% confidence interval of 0.97–1.00 and Gwet's AC1 of 0.98, indicating a very good level of inter-rater agreement. Thus, all items in the instrument were deemed statistically valid, and no items needed to be excluded.

3. Validation Decisions and Follow-Up Actions

Table 5. Validation Results and Follow-Up Recommendations

CVR Range	Category	Numbers of Items	Follow-up Action
$CVR=1.00$	Highly valid	46	Retained without substantial revision.
$CVR\geq0.80$	Valid	4	Revised based on qualitative feedback from panelists.
$CVR<0.80$	Invalid	0	Item discarded/deleted.

Achieving strong content validity in this study is a fundamental step in ensuring that each item comprehensively represents the theoretical construct of instructional media. This content validity serves as a key prerequisite before the instrument is used to assess the suitability of the media during the field trial phase. Specifically, quality assurance by experts ensures that the aspects of media, content, critical thinking, language, and design are theoretically sound before being implemented with students. Thus, the results of this CVR analysis provide a strong scientific basis that the instrument is ready to be used as a reliable evaluation tool in the development of AR-based physics learning media.

To test the stability of the CVR estimate with respect to a limited number of panelists (N=10), a sensitivity analysis was conducted. The test results showed that the CVR value remained stable above the critical threshold (≥0.80) even when a random reduction in the number of panelists was simulated. This proves that the inter-rater agreement in this instrument has strong internal consistency and is not sensitive to fluctuations in individual ratings. All items met the minimum statistical criteria (CVR ≥ 0.80), but 4 items (8%) were at the critical threshold. This indicates that although numerically valid, these items require minor revisions to improve their quality.

Discussion

1. Interpretation of Content Validity Results

The results of the Content Validity Ratio (CVR) analysis indicate that the AR flashcard validation instrument has very high content quality. A total of 46 out of 50 items (92%) received a CVR score of 1.00, reflecting unanimous agreement among the experts, while the remaining four items (M3, M7, MT7, CTS8) had a score of 0.80. All items exceeded the minimum critical threshold for the expert panel, making this instrument statistically suitable for use. Items with a score of 0.80 were retained with the note that minor revisions to the wording are needed to improve consistency of interpretation among raters.

The summary by aspect in Table 4 confirms the consistency of validity across all dimensions, with a total average CVR of 0.98 and a validity percentage of 100%. The Language and Design aspect achieved a perfect score (CVR = 1.00), indicating that this medium meets the required standards of readability and visual aesthetics for AR-based physics learning. Meanwhile, the Media, Content, and Critical Thinking Skills aspects, which exhibit a range of values (0.80–1.00), remain in the highly valid category, confirming that the instrument's theoretical constructs have been operationalized in a manner relevant to the learning objectives of Fluid Dynamics.

The extremely high CVR scores seen at the item and dimension level can be theoretically explained by the Cognitive Load Theory (CLT) and Constructivism. The positive scores for Language and Design suggest that the AR flashcards helped to lessen unnecessary cognitive load. The media has high readability, clear information hierarchy, and aesthetic visual design. The media does not overload the students' working memory with graphical distractions, so the students have cognitive resources to process the essential fluid dynamics concepts. Furthermore, the high validity in the Critical Thinking Skills (CTS) dimension aligns with constructivist principles, confirming that the media is designed to actively stimulate higher-order cognitive processes such as analysis and synthesis, rather than passive rote memorization, which is crucial for mastering abstract physics concepts.

Overall, Overall, the high CVR scores at both the item and dimension levels provide empirical evidence that this instrument possesses strong content validity prior to further testing. These findings align with modern validation principles, which emphasize the importance of expert consensus as the foundation for the quality of educational media evaluation instruments (Slater & Hasson, 2025). Thus, this AR flashcard validation instrument is deemed suitable for assessing the quality of learning materials, with recommendations for iterative improvements to four specific items to optimize the clarity of the assessment indicators.

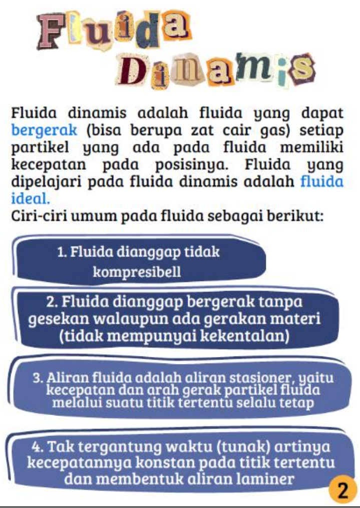
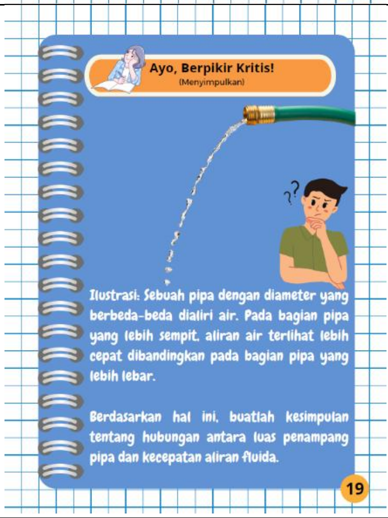
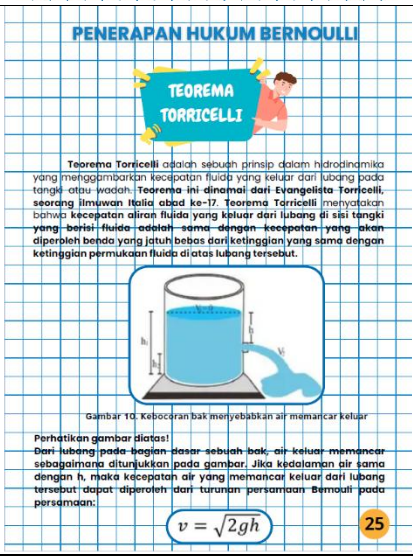
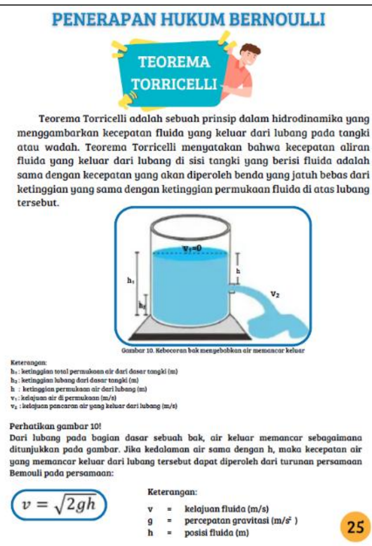
These results are an important extension to previous work in the field. Although studies such as (Dewi, P.S., & Kuswanto, 2023) have shown the effectiveness of interactive science media in improving analytical skills, the validation has been mainly based on descriptive expert judgements. The present study advances the field by providing rigorous statistical evidence with Gwet's AC1 (0.98). In this context, the inclusion of Gwet's AC1 is methodologically superior, as it corrects for prevalence bias, and demonstrates that the high inter-rater agreement is a true reflection of item essentiality, rather than a statistical artefact. Moreover, different from the other studies of the development of physics media in Indonesia, where only average Likert scores are used without critical statistical limits (Ridwan et al., 2021; Zavira & Sari, 2023), the bootstrap sensitivity analysis in this study proves that the validity estimates of this study are robust. The sensitivity analysis showed that the CVR values were still above the 0.80 threshold even when simulated with a smaller panel size, indicating that the high agreement is a true reflection of the quality of the instrument, not an artefact of a small, homogenous group of experts (Slater & Hasson, 2025).

The high level of panelist agreement (CVR = 0.98; Gwet's AC1 = 0.98) was interpreted with a critical eye toward potential biases such as the halo effect or social desirability, which were minimized through the use of forced-choice dichotomous scales and probabilistic critical thresholds (Ayre & Scally, 2014). Although the number of panelists (N=10) represents the minimum threshold recommended by Lawshe, the consistency of AC1 values and the results of the sensitivity analysis indicate that validity estimates remain robust. The four items that reached the threshold value (0.80) reflect the complexity of operationalizing AR features, not invalidity; therefore, they were revised minorly based on qualitative input coded thematically (frequency-weighted thematic matrix) into the visual, conceptual, and linguistic domains.

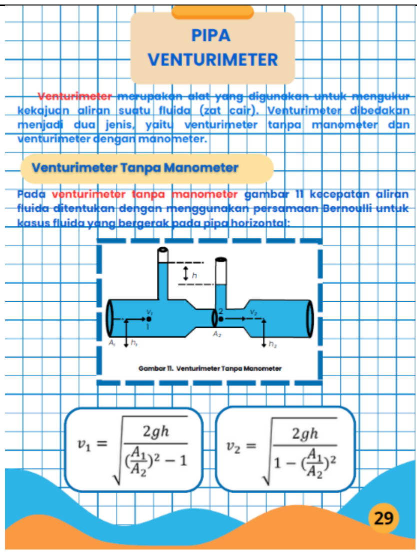
2. Integration of Qualitative Feedback and Product Revisions

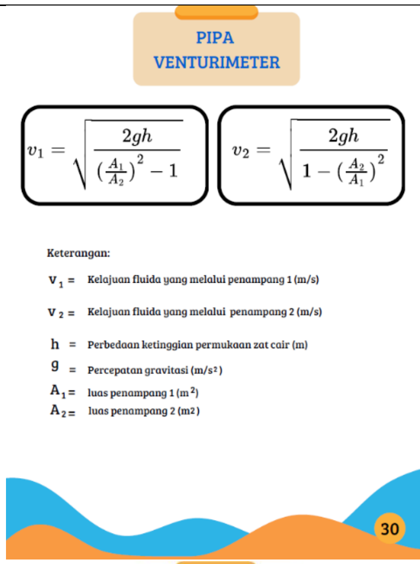
In addition to quantitative assessments, experts also provided qualitative feedback to improve the visual quality and readability of the materials. This feedback includes: (1) removing background lines from the media design to reduce visual distractions, (2) improving the image on page 19 to more clearly illustrate the relationship between velocity and cross-sectional area, and (3) adding equation labels on pages 25 and 29–30 to help students understand the mathematical formulations of fluid concepts. All of these suggestions were addressed through the first-stage revision of the product, in which design elements were simplified, illustrations were replaced with more explicit visuals, and equation annotations were added in accordance with the context of the material.

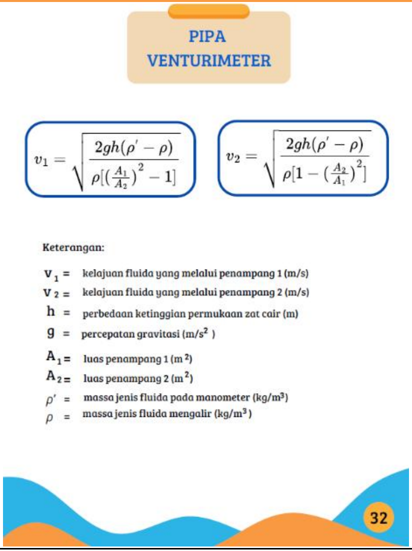
Table 6. Summary of Qualitative Feedback and Revision Follow-Up

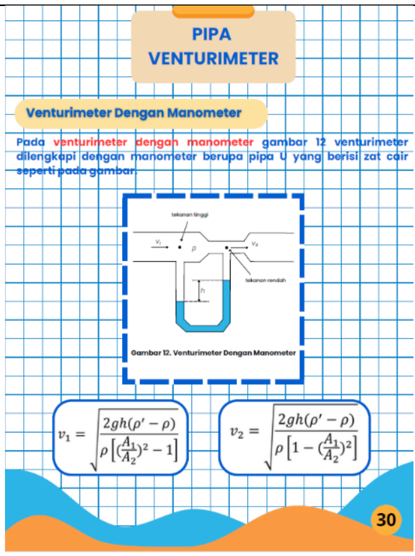
Comments	Before correction	After correction
Media design, remove the background lines		
Page 19: Correct the graph showing the relationship between velocity and cross-sectional area		
Page 25: Label the equations		

Pages 29–30:
Explain the equation









Augmented reality of an ideal fluid, with an explanation		Includes an additional explanation regarding ideal fluids
Augmented reality applied to turbulent flow in a pipe is introduced into the water		

Although the sample size (N=10) met Lawshe's statistical standards, future studies with a larger, cross-regional sample are recommended to strengthen the external validity and generalizability of the validation results.

CONCLUSION

Based on the results of the content validity analysis using an integrated quantitative-qualitative approach, it can be concluded that the validation instrument for Augmented Reality (AR) flashcards on the topic of Fluid Dynamics has met strict construct validity standards and is suitable for use as an evaluation tool. The five-dimensional framework developed not only ensures content accuracy but also pedagogical and cognitive alignment, making it suitable for adoption or adaptation by AR media developers, curriculum designers, and science education researchers for other materials or technological platforms. As a foundational step, this content validation should be followed by empirical field testing to establish the construct validity and reliability of the instrument, for example through ordinal reliability analysis (Cronbach's α or Rasch modeling). Future research is also recommended to involve multi-school or longitudinal designs to strengthen the generalizability of results and monitor the long-term impact of the media on students' conceptual understanding and critical thinking skills. Thus, this validation framework can serve as a transparent and replicable methodological standard in the development of immersive technology-based learning media grounded in empirical evidence.

ACKNOWLEDGMENTS

The researchers would like to thank all the validators for their contributions and feedback. We would also like to express our gratitude to Dr. Hamdi Akhsan, M.Si., and Dr. Sardianto Markos Siahaan, M.Pd., M.Si., for their support throughout the research process and the writing of this article.

REFERENCES

Afifah Aulia Zayrin, Hayatun Nupus, Khalista Khansa Maizia, Siska Marsela, Rully Hidayatullah, & Harmonedi, H. (2025). Analisis Instrumen Penelitian Pendidikan (Uji Validitas Dan Relibilitas Instrumen Penelitian). *Jurnal QOSIM Jurnal Pendidikan Sosial & Humaniora*, *3*(2), 780–789. https://doi.org/10.61104/jq.v3i2.1070

Amelia, M., & Nasution, B. (2025). Pengembangan Bahan Ajar Berbasis Video Powtoon pada Materi Gelombang Bunyi di Sma Kelas XI. *Jurnal Pendidikan Fisika*, *13*(2), 162–168. https://doi.org/10.24114/jpf.v13i2.60914

Arifta Nurjanah, Endah Retnowati; Augmented reality in the perspective of cognitive load theory. AIP Conf. Proc. 29 April 2024; 2622 (1): 090002. https://doi.org/10.1063/5.0134796

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, *47*(1), 79–86. https://doi.org/10.1177/0748175613513808

Bhakti, Y. B., Arthur, R., & Supriyati, Y. (2023). Development of an assessment instrument for critical thinking skills in Physics: A systematic literature review. *Journal of Physics: Conference Series*, *2596*(1). https://doi.org/10.1088/1742-6596/2596/1/012067

Dewi, P.S., & Kuswanto, H. (2023). The effectiveness of the use of augmented reality-assisted physics e-module based on pedicab to improve mathematical communication and critical thinking abilities. Journal of Technology and Science Education, 13(1), 53-64. https://doi.org/10.3926/jotse.1714

Erlangga, S. Y., Setyawan, D. N., Wijayanti, A., Winingsih, P. H., & Poort, E. A. (2024). Enhancing Students' Critical Thinking in Thermodynamics through Long Bumbung Local Wisdom-Based Physics Comics. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, *13*(1), 69. https://doi.org/10.24042/jipfalbiruni.v13i1.17872

López-Belmonte, J., Pozo-Sánchez, S., Lampropoulos, G., & Moreno-Guerrero, A. J. (2022). Design and validation of a questionnaire for the evaluation of educational experiences in the metaverse in Spanish students (METAEDU). *Heliyon*, *8*(11). https://doi.org/10.1016/j.heliyon.2022.e11364

Markos Siahaan, S., Astra Patriot, E., Scristia, S., Helen, H., Kalsum, U., & Elsa Putri, R. (2023). The Validity and Practicality of Augmented Reality-Based Media Development on Science Matter. *Jurnal Pendidikan MIPA*, *24*(4), 717–729. https://doi.org/10.23960/jpmipa/v24i4.pp717-729

Nabilah, N., Sari, S. D. P., & Sari, Y. M. (2025). Developing and Validating MCA-Equivalent Summative Assessment in Combinatorics Using Aiken's V Method. *Edumatika : Jurnal Riset Pendidikan Matematika*, *8*(2), 114–128. https://doi.org/10.32939/ejrpm.v8i2.5222

Nisrina, N., Rahmawati, I., & Hikmah, F. N. (2022). Pengembangan Instrumen Validasi Produk Multimedia Pembelajaran Fisika. *Lensa: Jurnal Kependidikan Fisika*, *10*(1), 32. https://doi.org/10.33394/j-lkf.v10i1.5278

Ridwan, M. R., Istiyono, E., & Widihastuti, W. (2021). Test Items Analysis of Mathematical Problem Solving Ability using a Classical Test Theory Approach. *Jurnal Pendidikan MIPA*, *22*(1), 98–111. https://doi.org/10.23960/jpmipa/v22i1.pp98-111

Romero Jeldres, M., Díaz Costa, E., & Faouzi Nadim, T. (2023). A review of Lawshe's method for calculating content validity in the social sciences. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1271335

Roy, R., Sukumar, G. M., Philip, M., & Gopalakrishna, G. (2023). Face, content, criterion and construct validity assessment of a newly developed tool to assess and classify work–related stress (TAWS– 16). *PLoS ONE*, *18*(1 January). https://doi.org/10.1371/journal.pone.0280189

Sarnita, F., Budi Adnyana, P., Rapi, N. K., Taman, S., & Bima, S. (2025). Uji Validitas dan Reliabelitas Instrumen Literasi Numerasi Materi Tata Surya untuk Siswa Tunanetra. *Jurnal Pendidikan Ilmu Pengetahuan Alam (JP-IPA)*, (02). https://doi.org/10.56842/jp-ipa

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)*, *10*(1), 50–63. https://doi.org/10.21831/reid.v10i1.63483

Slater, P., & Hasson, F. (2025). Quantitative Research Designs, Hierarchy of Evidence and Validity. *Journal of Psychiatric and Mental Health Nursing*, *32*(3), 656–660. https://doi.org/10.1111/jpm.13135

Tiara Ganish, S., Subali Jurusan Fisika, B., & Matematika dan Ilmu Pengetahuan Alam, F. (2025). Pengembangan Media Pembelajaran Fisika Materi Fluida Dinamis Berbasis Scratch untuk Meningkatkan Pemahaman Konsep Peserta Didik. In *Unnes Physics Education Journal Terakreditasi SINTA* (Vol. 14, Number 3). http://journal.unnes.ac.id/sju/index.php/upej

Wahyu Prananto, I., Rakhmawati, Y., Pamungkas, T., Negeri Yogyakarta, U., & Riau Kepulauan, U. (2022). Content Validity Ratio (CVR), Content Validity Index (CVI), and Confirmatory Factor Analysis (CFA) in Mathematics Learning Independence Instruments. *Kontinu: Jurnal Penelitian Didaktik Matematika*, *6*. https://doi.org/10.30659/kontinu.6.2.116-132

Yusuf, A., Suardana, I. N., & Selamet, K. (2021). Pengembangan Media Pembelajaran Flashcard IPA SMP Materi Tata Surya. *Jurnal Pendidikan Dan Pembelajaran Sains*, *4*(1), 12.

Zavira, N., & Sari, S. Y. (2023). Validitas Instrumen Aktivitas Belajar Siswa dalam Pembelajaran Fisika. *ARZUSIN*, *3*(6), 737–749. https://doi.org/10.58578/arzusin.v3i6.1940