Indonesian Journal of Educational Development (IJED)
Volume 7, Issue 1, 2026, pp. 82-95
ISSN: 2722-1059 (Online); ISSN: 2722-3671 (Print)
DOI: https://doi.org/10.59672/ijed.v7i1.6343

Development and validation of multimedia-based learning product instruments: SEM and Rasch model approaches

Sofyan*)[1], Erisa Kurniati[2], I Wayan Widana[3], Sabariah[4], Mazni binti Muhammad[5]
[1]Universitas Jambi, Jambi, Indonesia; sofyanzaibaski@unja.ac.id
[2]Universitas Jambi, Jambi, Indonesia; erisa.kurniati@unja.ac.id
[3]Universitas PGRI Mahadewa Indonesia, Bali, Indonesia; iwayanwidana@mahadewa.ac.id
[4]Universitas PGRI Adi Buana, Surabaya, Indonesia; sabariah@unipasby.ac.id
[5]Universiti Islam Melaka, Melaka, Malaysia; dr.mazni@unimel.edu.my
*)Corresponding author: Sofyan; E-mail addresses: sofyanzaibaski@unja.ac.id

Article Info

Article history:
Received April 22, 2026
Revised May 03, 2026
Accepted May 10, 2026
Available online May 20, 2026

Keywords: Instrument validation, Learning evaluation, Multimedia-based learning, Rasch model, SEM

Abstract. Despite the rapid expansion of multimedia-based learning in education, the availability of psychometrically validated evaluation instruments remains limited, underscoring the urgent need for reliable, standardized measurement tools. This study aims to develop and validate a multimedia-based learning product evaluation instrument that measures three main constructs: feasibility, practicality, and attractiveness. The study used an instrument development design with an integrated psychometric approach that combines classical measurement theory (Classical Test Theory), Structural Equation Modeling (SEM), and the Rasch Model. Content validity was assessed using expert judgment, as measured by Aiken's V coefficient and the Content Validity Ratio (CVR). Construct validity was tested through Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). Structural analysis was conducted to examine the relationships between constructs, and multi-group analysis (teachers and students) was conducted to test model invariance. Data were collected from 257 respondents, comprising 153 students and 104 teachers. The results showed that the instrument had high content validity (mean Aiken's V = 0.89), very high reliability (α = 0.93), and a good measurement model fit (CFI = 0.96; RMSEA = 0.054). Rasch analysis showed item reliability of 0.96, and no serious misfit was found. Multi-group analysis showed the model was invariant between teachers and students. Thus, this instrument meets modern psychometric standards and is suitable for use in evaluating multimedia-based learning products.

Introduction

The digital transformation in education has shifted the learning paradigm from conventional models to multimedia-based learning and digital learning resources. The integration of multimedia, Learning Management Systems (LMS), interactive applications, and online learning resources is no longer a complement but has become a core component of modern instructional design (Widana & Ratnaya, 2021). Recent studies have shown that the systematic use of digital multimedia can increase learning engagement and effectiveness, and enhance flexibility in educational access (Bond et al., 2020; Martin et al., 2020; Cheng, 2020; Agustian et al, 2025). However, this increased use of learning technology has not always been accompanied by psychometrically validated evaluation instruments.

Ideally, evaluation of multimedia-based learning products should be conducted using instruments that are not only content-relevant but also empirically tested through rigorous construct validation and reliability procedures (Eka Setiawati, 2024). Comprehensive evaluations enable educational developers and practitioners to ensure that the products they use are truly feasible, practical, and engaging for users. However, in practice, many research and development (R&D) projects still use ad hoc evaluation instruments that have not undergone systematic content validity testing and have not been tested using modern measurement approaches such as Structural Equation Modeling (SEM) or Item Response Theory (IRT).

Conceptually, the quality of learning products must be viewed through the lens of systematic instructional design. The ADDIE model emphasizes that learning effectiveness is determined by the alignment between needs analysis, design, development, implementation, and evaluation (Branch, 2019). The systematic instructional design approach also emphasizes the importance of coherence between learning objectives, materials, strategies, and evaluation (Dick et al., 2015). Therefore, learning product evaluation instruments ideally reflect the dimensions of content appropriateness, pedagogical suitability, and multimedia and technical design quality in an integrated manner.

From a multimedia design perspective, the Cognitive Theory of Multimedia Learning explains that effective learning occurs when information is presented through a structured integration of text and visuals, thus facilitating the dual coding process in working memory (Mayer, 2014). This principle is reinforced by Cognitive Load Theory, which emphasizes that learning design must manage intrinsic, extrinsic, and germane cognitive load to avoid hindering the knowledge construction process (Sweller et al., 2011). Research over the past five years has shown that multimedia design that does not account for the principle of cognitive load can reduce the effectiveness of digital learning (Leppink, 2020; Chen et al., 2022).

In addition to design aspects, motivational dimensions are also important determinants of the success of multimedia-based learning. The ARCS (Attention, Relevance, Confidence, Satisfaction) model emphasizes that attractiveness, material relevance, self-confidence, and learning satisfaction contribute to learning engagement and persistence (Keller, 2010). Recent studies have shown that interactivity and usability elements in digital learning multimedia significantly increase student engagement (Bond et al., 2020; Bedenlier et al., 2021).

User acceptance of technology is also a key factor. The Technology Acceptance Model (TAM) states that perceived ease of use and usefulness influence attitudes and intentions toward technology use (Davis, 1989). Recent research on digital education reinforces the relevance of TAM for explaining the adoption of e-learning and cloud-based systems (Baber, 2021; Teo & Zhou, 2021; Cheng, 2020). Habibi et al. (2024) showed that access to and readiness for technology in vocational schools influence the optimal use of learning technology, indicating the importance of the practicality dimension in evaluating learning products.

In the context of instrument development, Sofyan et al. (2023) and Megawaty et al. (2025) developed the TPACK–UoTI instrument for elementary school teachers using a modern SEM-based validation procedure. This study emphasized the importance of comprehensive psychometric validation in the development of educational measurement tools. However, most instrument development research still focuses on a single dimension (e.g., teacher competence or technology acceptance). It has not integrated the dimensions of instructional feasibility, practicality of use, and motivational appeal into a unified evaluative framework.

Several recent studies have also shown a trend toward using SEM and the Rasch Model to develop educational instruments and improve measurement accuracy (Boateng et al., 2020; Taber, 2023; Koo & Li, 2021). However, there is still limited research combining content validity (Aiken's V and CVR), CFA, structural SEM, and the Rasch Model in a single development of multimedia-based learning product evaluation instruments. This methodological gap indicates a significant research gap.

Thus, the research gap in this study can be formulated as follows: (1) there are still limited multimedia-based learning product evaluation instruments developed through comprehensive psychometric validation procedures; (2) there are not many studies that integrate the perspectives of instructional design, multimedia theory, learning motivation, and technology acceptance in one evaluative model; and (3) there is minimal integration between SEM and the Rasch Model in the validation of multimedia learning evaluation instruments.

The novelty of this research lies in the development of a multimedia-based learning product evaluation instrument that integrates three main dimensions, namely: feasibility, practicality, and attractiveness, in an integrated conceptual framework based on the theory of instructional design, multimedia learning, ARCS, and TAM. In addition, this research combines Classical Test Theory and Item Response Theory approaches through Aiken's V analysis, CFA, SEM, multi-group SEM, and the Rasch Model to provide double validation at the construct and item levels. The urgency of this research is heightened by the accelerated digitalization of education and the disparity in the quality of multimedia learning implementation across school levels. Without robustly validated evaluation instruments, decision-making regarding the adoption and development of digital learning products can be subjective and inaccurate.

Therefore, this study aims to develop and validate a comprehensive multimedia-based learning product evaluation instrument using an integrated psychometric approach. To address this objective, the study is guided by the following research questions: (1) How can a multimedia-based learning evaluation instrument be systematically developed based on instructional design, multimedia learning theory, ARCS, and TAM frameworks? (2) To what extent does the developed instrument demonstrate strong content validity, construct validity, and reliability? (3) How do the constructs of feasibility, practicality, and attractiveness structurally influence learning effectiveness within the SEM framework? and (4) Does the measurement model demonstrate invariance between teacher and student groups? Based on these research questions, the study aims to develop a psychometrically sound instrument to comprehensively and empirically evaluate multimedia-based learning products.

Method

Research Design

This study employed an instrument-development design with a quantitative, psychometric-based approach. This design was chosen because developing educational measurement tools requires a systematic validation process to ensure adequate validity and reliability. According to DeVellis (2017), instrument development must go through the stages of construct conceptualization, item development, content validation, and empirical testing using factor analysis and reliability analyses. In this study, Classical Test Theory (CTT) was used to assess internal consistency. In contrast, Structural Equation Modeling (SEM) was used to assess construct validity and the relationships among latent variables (Hair et al., 2019). To strengthen the quality of measurement at the item level, this study also integrated the Rasch Model within Item Response Theory (IRT), which allows evaluation of item stability independent of sample characteristics (Bond & Fox, 2015). The integration of CTT, SEM, and Rasch provides a comprehensive methodological foundation for

developing multimedia-based learning evaluation instruments. The procedural model design for this study is shown in Image 1.

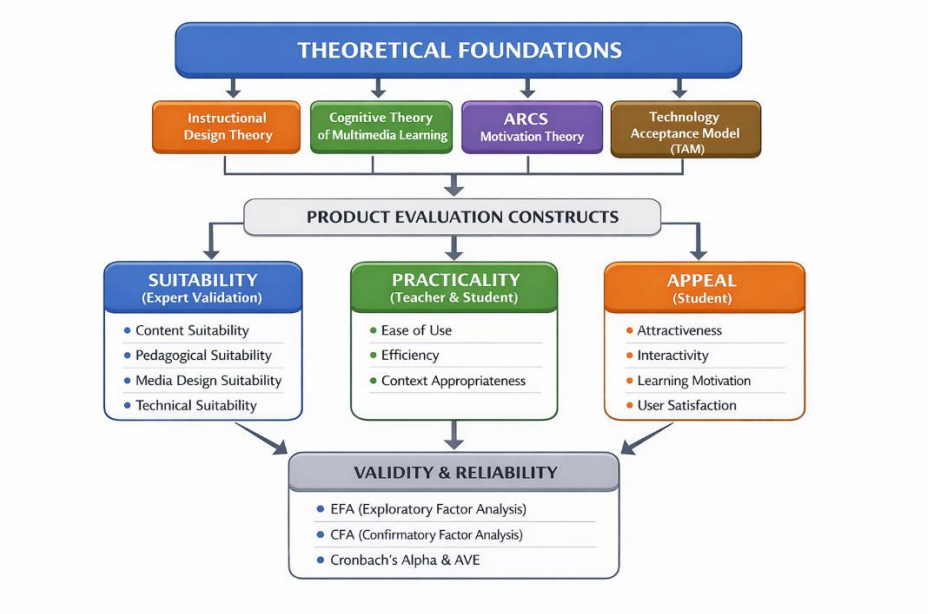

Image 1. Procedural Model of Instrument Development

Location and Time of Research

The research was conducted at several elementary, junior high, senior high, and vocational high schools in Jambi City. Locations were selected purposively, considering schools that had implemented multimedia-based learning and digital learning resources. Data collection took place in the odd semester of the 2025/2026 academic year, between August and November 2025. This period was chosen because the learning process had been stable, and respondents had sufficient experience using the evaluated learning products.

Participants

The study involved 257 respondents, comprising 153 students and 104 teachers. The student respondents came from junior high, senior high, and vocational high schools. The selection of students at the secondary education level was based on considerations of their cognitive development. During adolescence, individuals generally reach the formal operational stage, enabling abstract and reflective thinking about their learning experiences. With these skills, students are considered capable of providing rational assessments of the quality and experience of using multimedia-based learning products.

The teacher respondents came from various educational levels, namely elementary, middle, high, and vocational high schools. This diversity of levels was intended to obtain a broader professional perspective in assessing the pedagogical and technical aspects of learning products. Teachers' professional judgment is important in educational evaluation because they have direct experience in curriculum implementation and classroom learning practices. Furthermore, the variation in respondent characteristics allowed for multi-group analysis to test model invariance (Hair et al., 2019). The sample size of this study also met the minimum recommendation for Maximum Likelihood-based SEM analysis, which requires more than 200 respondents to produce stable parameter estimates (Hair et al., 2019).

In addition to the primary respondents, the draft instrument was initially validated by three experts in educational technology, multimedia learning, and language. These experts came from Ganesha

University of Education, the University of Lampung, and the University of Jambi. Each expert holds a doctoral degree and serves as a Professor in their respective field. The experts' assessments were used to evaluate content validity, wording clarity, and the suitability of the instrument's indicators prior to implementation.

Instrument Development

The feasibility dimension consists of four indicators: content feasibility, pedagogical feasibility, multimedia design feasibility, and technical feasibility. The practicality dimension includes ease of use, time efficiency, flexibility, and integration into learning, while the attractiveness dimension comprises attention, relevance, confidence, and satisfaction. These indicators were operationalized based on established frameworks relevant to instructional design, technology use, and learning motivation to ensure conceptual alignment and construct validity.

Data Analysis Procedure

The analysis was conducted in stages. First, content validity was tested using Aiken's V and Content Validity Ratio (CVR) to ensure item relevance to the construct (Aiken, 1985). Second, reliability was assessed using Cronbach's Alpha and the Intraclass Correlation Coefficient (ICC) to measure internal consistency and inter-rater agreement, respectively. Third, construct validity was tested using Confirmatory Factor Analysis (CFA) with criteria for factor loadings $\geqslant 0.50$, Average Variance Extracted (AVE) $\geqslant 0.50$, and Composite Reliability (CR) $\geqslant 0.70$ (Hair et al., 2019). Next, Structural Equation Modeling (SEM) analysis was conducted to examine the relationships between latent constructs. To strengthen measurement validity, Rasch analysis was used to evaluate item and person reliability and the separation index (Bond & Fox, 2015). Finally, a multi-group SEM analysis was conducted to test model invariance between teacher and student groups.

Ethical Considerations

This research has obtained approval and permission from the Research and Community Service Institute of the University of Jambi, the Jambi Provincial Education Office, and the Jambi City Education Office. All respondents were provided with information regarding the research objectives and guaranteed confidentiality. Data collection was conducted after obtaining informed consent from the participants and official permission from the school. The principles of anonymity, confidentiality, and use of data solely for academic purposes were maintained in accordance with ethical guidelines for educational research.

Results and Discussion

In addition to the three primary constructs (feasibility, practicality, and attractiveness), this study also incorporates learning effectiveness as an outcome variable to examine the impact of multimedia-based learning quality on learning outcomes. The inclusion of effectiveness is intended to provide a more comprehensive evaluation framework by linking product quality dimensions with learning performance.

Descriptive Statistics

Table 1. Descriptive Statistics of Constructs

Construct	Mean	Elementary School	Skewness	Kurtosis
Feasibility	4.21	0.53	-0.41	0.38
Practicality	4.18	0.57	-0.35	0.29
Attractiveness	4.25	0.51	-0.47	0.44
Effectiveness	4.17	0.60	-0.32	0.21

The descriptive statistics presented in Table 1 indicate that all measured constructs, feasibility, practicality, attractiveness, and effectiveness, have mean scores above 4.00, reflecting a generally high level of positive perception among respondents toward multimedia-based learning products. The relatively low standard deviations across all constructs suggest that respondents' answers are consistently clustered around the mean, indicating a stable and homogeneous pattern of responses. This consistency implies that the evaluated learning products are perceived similarly by participants in terms of both quality and usability. Such findings are important in educational measurement because they suggest that the instrument captures a shared perception among respondents, thereby strengthening the reliability of the descriptive results. In the context of multimedia learning evaluation, high mean values also indicate that the products being assessed are considered feasible, practical, attractive, and effective, aligning with the intended design and implementation goals of technology-enhanced learning.

Furthermore, the distributional properties of the data, as indicated by skewness and kurtosis values ranging between −1 and +1, confirm that the dataset follows a normal distribution. This is a critical prerequisite for conducting advanced statistical analyses such as Structural Equation Modeling (SEM), particularly when using Maximum Likelihood estimation techniques. According to Hair et al. (2019), normality assumptions must be met to ensure unbiased parameter estimates and accurate model fit evaluation in SEM. The absence of extreme skewness or kurtosis in the data suggests that there are no significant outliers or distributional distortions that could compromise the validity of subsequent analyses. Therefore, the data structure is statistically appropriate for further inferential testing, including CFA and SEM, which rely on the assumption of multivariate normality.

From an analytical perspective, these descriptive findings provide an important foundation for interpreting the subsequent validity and structural model results. The high mean scores across constructs suggest that respondents not only perceive the multimedia learning products positively but also experience usability and effectiveness benefits. This aligns with broader research in educational technology, which shows that well-designed digital learning environments tend to yield positive user perceptions and engagement. Moreover, the normal distribution and low response variability indicate that the instrument performs well at capturing consistent evaluations across respondents. As a result, the descriptive statistics not only confirm the quality of the data but also reinforce the credibility of the measurement instrument, supporting its use in more complex psychometric analyses and contributing to a comprehensive evaluation of multimedia-based learning products.

Content Validity and Reliability

The results of the content validity analysis showed that most items had a Content Validity Ratio (CVR) of 1.00, indicating unanimous agreement among experts on each item's relevance to the intended construct. This finding indicates a very high level of content representativeness and supports the instrument's adequacy at the conceptual level. In instrument development, content validity serves as a foundational step to ensure that each item accurately reflects the construct domain before proceeding to more complex statistical validation. According to Aiken (1985), expert judgment plays a critical role in assessing item relevance, clarity, and representativeness, particularly through indices such as Aiken's V and CVR. Furthermore, best practices in scale development emphasize that strong content validity is an essential prerequisite for ensuring construct validity and measurement accuracy in subsequent analyses (Boateng et al., 2020; Widana et al., 2021). This aligns with the present findings, which confirm that the high CVR values indicate the instrument items are theoretically grounded and aligned with the intended dimensions of feasibility, practicality, and attractiveness.

In addition, the high level of expert agreement observed in this study contributes significantly to the clarity and precision of measurement indicators. When experts consistently evaluate items as relevant, it indicates that the operational definitions of constructs have been clearly translated into measurable indicators. Zamanzadeh et al. (2020) highlight that strong agreement among experts enhances both item clarity and the representativeness of the construct domain, thereby reducing ambiguity in interpretation. This is particularly important in the context of multimedia-based learning evaluation, where constructs often integrate pedagogical, technological, and motivational dimensions. The integration of these dimensions requires carefully designed indicators that reflect theoretical coherence, as emphasized in systematic instructional design models such as ADDIE (Branch, 2019) and the structured alignment of learning components (Dick et al., 2015). Therefore, the high CVR results not only indicate agreement but also suggest that the instrument successfully captures the multidimensional nature of multimedia learning evaluation.

From the reliability perspective, the Cronbach's Alpha coefficient of 0.93 indicates excellent internal consistency among the instrument items. This suggests that the items within each construct consistently and stably measure the same underlying concept. According to Taber (2023), reliability coefficients above 0.90 are considered highly satisfactory and reflect strong construct stability, especially in complex measurement contexts such as digital and technology-enhanced learning environments. High internal consistency is essential for minimizing measurement error and ensuring the instrument produces dependable results across different samples. Moreover, DeVellis (2017) emphasizes that a high Cronbach's Alpha indicates homogeneity among items, which strengthens the credibility of the scale as a reliable measurement tool. In the context of this study, the high reliability coefficient confirms that the developed instrument is robust and suitable for evaluating multimedia-based learning products.

Furthermore, the Intraclass Correlation Coefficient (ICC) value of 0.91 demonstrates a very high level of agreement among expert raters, reinforcing the reliability of the validation process. According to Koo and Li (2021), ICC values above 0.90 indicate "excellent agreement," suggesting that ratings from different experts are highly consistent and reliable. This level of agreement is crucial to ensuring that the validation results are not influenced by subjective bias or rater variability. The combination of strong content validity (CVR = 1.00), high internal consistency (α = 0.93), and excellent inter-rater agreement (ICC = 0.91) demonstrates that the instrument meets rigorous psychometric standards. These findings are consistent with contemporary trends in educational measurement, which advocate integrating expert validation and advanced statistical techniques to enhance the quality of technology-based assessment instruments (Boateng et al., 2020; Hair et al., 2019). Consequently, the instrument developed in this study is both conceptually sound and empirically reliable, supporting its use in evaluating multimedia-based learning products across diverse educational contexts.

Construct Validity (CFA)

Table 2. Measurement Model (CFA) Results

Construct	Loading	AVE	CR	Alpha
Feasibility	0.64–0.88	0.68	0.94	0.93
Practicality	0.62–0.86	0.65	0.92	0.91
Attractiveness	0.66–0.90	0.71	0.95	0.94
Effectiveness	0.60–0.85	0.63	0.90	0.89

The results of the Confirmatory Factor Analysis (CFA) indicate that the measurement model meets the criteria for both convergent validity and construct reliability. The model fit indices indicate a strong fit to the empirical data, with CFI = 0.96, TLI = 0.95, RMSEA = 0.054, and SRMR = 0.046.

These values fall within the recommended thresholds for a well-fitting model, confirming that the hypothesized factor structure is consistent with the observed data. According to Hair et al. (2019), model fit indices such as CFI and TLI values above 0.90, along with RMSEA values below 0.08, indicate an acceptable-to-good model fit. Therefore, the CFA results provide strong empirical evidence that the measurement model used in this study is statistically sound and appropriate for representing the underlying constructs.

Furthermore, all constructs in the model achieved Average Variance Extracted (AVE) values greater than 0.50 and Composite Reliability (CR) values exceeding 0.90, indicating strong convergent validity and high construct reliability. AVE values above 0.50 suggest that more than half of the variance in the indicators is explained by the latent construct. In contrast, high CR values indicate consistency among indicators in measuring the same construct (Hair et al., 2019). These results confirm that each construct's feasibility, practicality, and attractiveness are measured with high precision and internal coherence. Such findings are essential in instrument development, as they ensure that the measurement model captures the theoretical constructs accurately and minimizes measurement error.

In addition, the CFA results reveal that all indicators have strong, significant factor loadings, further supporting convergent validity. High factor loadings indicate that each indicator contributes meaningfully to its respective latent variable, reinforcing the unidimensionality of each construct. This aligns with the principles of systematic measurement development, where indicators must be closely aligned with the conceptual definitions of constructs (DeVellis, 2017). The strong relationship between indicators and latent variables suggests that the instrument has successfully operationalized abstract constructs into measurable variables. Consequently, the instrument demonstrates both theoretical coherence and empirical robustness, which are essential characteristics of high-quality educational measurement tools.

From a theoretical perspective, these findings are consistent with the principles of systematic instructional design, which emphasize the alignment between learning objectives, instructional strategies, and evaluation measures. Models such as ADDIE proposed by Branch (2019) and the systematic design framework by Dick et al. (2015) highlight the importance of coherence between instructional components and assessment instruments. The good model fit achieved in this study indicates that the developed instrument aligns well with these theoretical foundations. As a result, the constructs of feasibility, practicality, and attractiveness are not only statistically valid but also conceptually grounded, thereby strengthening their legitimacy as key dimensions for evaluating multimedia-based learning products in contemporary digital education contexts.

Structural SEM

Table 3. Structural Model Results

Track	β	t	p
F → P	0.74	12.33	<.001
P → A	0.68	9.71	<.001
A → E	0.71	8.89	<.001
F → E	0.29	2.53	.012

The Structural Equation Modeling (SEM) results presented in Table 3 indicate that all hypothesized relationships are statistically significant and support the proposed research model. Specifically, feasibility has a strong, significant effect on practicality ($\beta = 0.74$, $p < 0.001$); practicality significantly influences attractiveness ($\beta = 0.68$, $p < 0.001$); and attractiveness significantly affects effectiveness ($\beta = 0.71$, $p < 0.001$). Additionally, feasibility also shows a direct, though relatively

weaker, effect on effectiveness (β = 0.29, p = 0.012). These findings suggest that the structural relationships among the constructs are both meaningful and theoretically coherent. According to Hair et al. (2019), significant path coefficients in SEM indicate that the empirical data support the hypothesized causal relationships between latent variables. Therefore, the model demonstrates strong explanatory power in describing how different dimensions of multimedia-based learning evaluation interact.

From a theoretical standpoint, the relationship among feasibility, practicality, attractiveness, and effectiveness reflects a coherent progression of learning-quality dimensions. Feasibility serves as the foundational condition that enables practical use, while practicality enhances user interaction and engagement, ultimately contributing to attractiveness and learning effectiveness. Rather than merely confirming existing theoretical models, these findings provide empirical evidence that integrates usability, motivation, and instructional quality into a unified structural framework. This indicates that the effectiveness of multimedia learning is not determined by a single factor but emerges from the interaction between design quality, usability, and learner engagement (Baber, 2021; Teo & Zhou, 2021).

Furthermore, the significant relationship between attractiveness and effectiveness reinforces the relevance of motivational and cognitive learning theories in explaining learning outcomes. The ARCS model proposed by Keller (2010) highlights that attention, relevance, confidence, and satisfaction are key drivers of learning motivation, which directly impacts learning performance. In parallel, the Cognitive Theory of Multimedia Learning, developed by Mayer (2014), explains that well-designed multimedia facilitates cognitive processing by effectively integrating visual and verbal information. When multimedia learning products are both attractive and well-structured, they enhance learners' cognitive engagement and knowledge construction processes. Thus, the structural model in this study not only demonstrates statistical validity but also integrates instructional design principles, motivational theory, and cognitive theory into a unified framework, strengthening its contribution to the evaluation of multimedia-based learning in contemporary digital education contexts.

Multi-group SEM

Table 4. Comparison of Coefficients Between Groups

Track	Teacher β	Student β	Δβ	Significant
F → P	0.71	0.77	0.06	No
P → A	0.59	0.73	0.14	Yes
A → E	0.74	0.69	0.05	No

The results of the multi-group Structural Equation Modeling (SEM) analysis presented in Table 4 reveal important insights into the structural relationships across different respondent groups, namely teachers and students. Overall, most structural paths demonstrate invariance between the two groups, indicating that the measurement and structural model is generally stable across populations. Specifically, the relationships between feasibility and practicality (F → P) and between attractiveness and effectiveness (A → E) show only minor differences in path coefficients and are not statistically significant. This suggests that both teachers and students share similar perceptions regarding the foundational role of feasibility in shaping practicality and the influence of attractiveness on learning effectiveness. According to Hair et al. (2019), multi-group invariance testing is essential to ensure that a model operates consistently across different groups, thereby enhancing the generalizability and robustness of the findings.

However, a notable exception is observed in the relationship between practicality and attractiveness (P → A), where a significant difference is found between teachers ($\beta = 0.59$) and students ($\beta = 0.73$), with a statistically significant difference ($\Delta\beta = 0.14$). This indicates that practicality has a stronger influence on perceived attractiveness among students compared to teachers. In other words, students tend to place greater importance on ease of use, efficiency, and usability when evaluating the attractiveness of multimedia-based learning products. This difference highlights the distinct perspectives of end users (students) and professional evaluators (teachers), with students more directly engaged with the usability aspects of learning technologies in their daily learning experiences.

From a theoretical perspective, this finding can be explained through the lens of the Technology Acceptance Model (TAM) developed by Davis (1989). TAM posits that perceived ease of use plays a critical role in shaping users' attitudes toward technology. Students, as primary users of multimedia learning systems, are more sensitive to usability factors because these directly affect their interaction experience and learning efficiency. In contrast, teachers may evaluate learning products more holistically, considering pedagogical alignment, content quality, and instructional effectiveness in addition to usability. This interpretation is supported by studies in digital education, which indicate that user experience and system usability are key determinants of engagement and satisfaction, particularly among students (Baber, 2021; Teo & Zhou, 2021).

Furthermore, this finding is consistent with research on student engagement with educational technology, which emphasizes that perceived usability significantly influences learners' motivation and engagement. Studies by Michael Bond et al. (2020) show that interactive and user-friendly learning environments enhance student engagement more strongly than other design factors. Students' greater sensitivity to practicality reflects their need for intuitive, efficient learning tools that reduce cognitive effort and enable seamless interaction. This can also be interpreted through the lens of cognitive load theory, in which simpler, more usable systems help minimize extraneous cognitive load, thereby improving learning experiences (Sweller et al., 2011; Leppink, 2020). Therefore, the multi-group SEM results not only confirm the structural stability of the model but also provide valuable insights into stakeholder differences, enriching the literature on multimedia learning evaluation and emphasizing the importance of user-centered design in educational technology development.

Rasch Model

Table 5. Rasch Analysis Results

Indicator	Mark
Item Reliability	0.96
Person Reliability	0.92
Item Separation	4.85
Person Separation	3.41
Misfit Items	6 (minor revision)

The Rasch analysis results presented in Table 5 indicate a very high level of measurement stability and precision. The item reliability coefficient of 0.96 and the person reliability of 0.92 demonstrate that the instrument performs consistently in both item functioning and respondent consistency. Additionally, the item separation index of 4.85 and person separation index of 3.41 indicate a strong ability of the instrument to distinguish between different levels of item difficulty and respondent ability. These values suggest that the instrument has excellent discriminative power, allowing it to effectively capture variations in respondents' perceptions of multimedia-based learning products.

Such findings confirm that the measurement system is robust and can produce stable, reliable estimates across respondent groups.

Furthermore, the identification of only six items with minor misfit indicates that the instrument's overall quality is high and requires only minimal revisions. In Rasch measurement, misfit items are those that do not align well with the expected measurement model, potentially introducing noise into the data. However, the small number of misfit items found in this study suggests that most items function as intended and contribute meaningfully to the measurement construct. This reinforces the conclusion that the instrument is well-calibrated and suitable for practical application, particularly in educational contexts where measurement accuracy is critical.

From a theoretical and methodological perspective, the Rasch Model provides a significant advantage in evaluating item quality independently of the sample. As highlighted by Bond & Fox (2015), the Rasch approach allows researchers to assess item characteristics such as difficulty and fit without being overly influenced by specific sample distributions. This property enhances the instrument's generalizability and ensures that measurement results remain stable across different populations. The high separation indices observed in this study further indicate that the instrument can effectively differentiate respondents by their perceptions, thereby strengthening its diagnostic capability for evaluating multimedia learning products.

Moreover, integrating Rasch analysis with Structural Equation Modeling (SEM) offers a comprehensive approach to instrument validation. While SEM focuses on validating relationships between latent constructs and ensuring model fit at the structural level, Rasch analysis evaluates the quality of individual items at the measurement level. According to Hair et al. (2019), SEM is highly effective for testing theoretical models and relationships among constructs. In contrast, Rasch analysis ensures that each item contributes meaningfully to the measurement scale. The combination of these two approaches provides dual validation at both the construct and item levels, thereby significantly enhancing the instrument's overall credibility and rigor.

This study also offers important theoretical contributions by integrating multiple perspectives, including instructional design, multimedia learning, motivation, and technology acceptance, into a unified evaluative framework. The findings reinforce the view that instructional feasibility serves as the foundational dimension influencing both practicality and attractiveness. This is consistent with systematic instructional design principles proposed by Branch (2019) and by Dick et al. (2015), which emphasize alignment among objectives, instructional strategies, and evaluation. Such alignment ensures that learning products are not only technically sound but also pedagogically effective.

In addition, the significant relationship between practicality and attractiveness extends the interpretation of the Technology Acceptance Model (TAM) developed by Davis (1989). The findings demonstrate that perceived ease of use influences not only behavioral intention but also affective responses such as interest and motivation. This aligns with the ARCS motivational model introduced by Keller (2010), which highlights the importance of attention, relevance, confidence, and satisfaction in driving learner engagement. Thus, the instrument captures not only the functional aspects of multimedia learning but also the motivational dimensions that are critical to effective learning outcomes.

From a practical perspective, these findings highlight the importance of designing multimedia learning products that are not only technically feasible but also easy to use and engaging for learners. Educational institutions and developers can utilize this instrument as a standardized evaluation tool to assess product quality and guide continuous improvement. In particular, the

strong relationship between practicality and attractiveness suggests that user-friendly design should be prioritized to enhance learner engagement and optimize learning outcomes. This aligns with the principles of Cognitive Load Theory, which emphasize reducing extraneous cognitive load and promoting clear, coherent multimedia presentation (Sweller et al., 2011; Mayer, 2014). These insights have direct implications for instructional designers, teachers, and policymakers in ensuring the effective implementation of digital learning technologies.

Conclusion

This study developed and validated a multimedia-based learning product evaluation instrument using a comprehensive psychometric approach that integrates Classical Test Theory, Structural Equation Modeling (SEM), and the Rasch Model. The findings indicate that the instrument demonstrates strong content validity, high reliability, and robust construct validity, with all constructs empirically supported and structurally related within a coherent measurement framework. The analysis confirms that feasibility, practicality, and attractiveness function as key dimensions influencing learning effectiveness, while the integration of CFA, SEM, and Rasch analysis provides complementary validation at both construct and item levels, thereby enhancing measurement precision and overall instrument robustness. In addition, the instrument exhibits stability across teacher and student groups, indicating its applicability across different educational stakeholders and contexts. From a practical perspective, the results underscore the importance of using standardized, psychometrically validated instruments to evaluate multimedia-based learning products, particularly amid rapid digital transformation in education, where many innovations are implemented without rigorous evaluation. The availability of this validated instrument offers a reliable tool for educators, developers, and policymakers to support evidence-based decision-making, guide the design and improvement of learning products, and ensure alignment with pedagogical objectives and user needs. Furthermore, this study contributes to the advancement of educational measurement by providing an integrated validation framework that combines multiple analytical approaches, which can serve as a reference for future research and development in educational technology. Future studies are recommended to further examine the applicability of this instrument across diverse cultural, institutional, and regional contexts to strengthen its generalizability and support the development of globally relevant standards for evaluating multimedia-based learning products.

Acknowledgements

The authors would like to thank the Research and Community Service Institute of Universitas Jambi, the teachers and students of junior high, senior high, and vocational high schools within the Jambi Provincial Education Office, and the Jambi City Education Office for their participation in this research. Thanks are also extended to the validation partners from Universitas Jambi, Universitas Lampung, and Universitas Pendidikan Ganesha. Furthermore, thanks are extended to the research and writing partners from Universitas PGRI Mahadewa Indonesia, Universitas PGRI Adi Buana Surabaya, and the Islamic University of Melaka, Malaysia. Hopefully, the results of this research will contribute to the development of education and learning in Indonesia and the ASEAN region.

Bibliography

Agustian, M., Jerica, Ernestya, T. G., & Ekarina. (2025). Development of interactive multimedia based on cultural literacy for Pancasila education in elementary schools. *Indonesian Journal of Educational Development (IJED)*, *6*(3), 1085–1101. https://doi.org/10.59672/ijed.v6i3.5349

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement, 45*(1), 131–142. https://doi.org/10.1177/0013164485451012
Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology. *International Journal of Educational Technology in Higher Education, 17*(2). https://doi.org/10.1186/s41239-019-0176-8
Bond, T. G., & Fox, C. M. (2015). *Applying the rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
Branch, R. M. (2019). *Instructional design: The ADDIE approach*. Springer. https://doi.org/10.1007/978-0-387-09506-6
Baber, H. (2021). Modelling the acceptance of e-learning during the COVID-19 pandemic. *The International Journal of Management Education, 19*(2), 100503. https://doi.org/10.1016/j.ijme.2021.100503
Bedenlier, S., Bond, M., Buntins, K., & Zawacki-Richter, O. (2021). Facilitating student engagement through educational technology. *Educational Technology Research and Development, 69*, 165–190. https://doi.org/10.1007/s11423-020-09873-4
Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2020). Best practices for developing and validating scales. *Frontiers in Public Health, 8*, 149. https://doi.org/10.3389/fpubh.2020.00149
Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model*. Routledge.
Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology. *International Journal of Educational Technology in Higher Education, 17*(2). https://doi.org/10.1186/s41239-019-0176-8
Cheng, Y.-M. (2020). Students' satisfaction and continuance intention. *Computers in Human Behavior, 107*, 105993. https://doi.org/10.1016/j.chb.2019.105993
Chen, O., et al. (2022). Multimultimedia design and cognitive load. *Computers & Education, 180*, 104435. https://doi.org/10.1016/j.compedu.2022.104435
Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008
DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage Publications.
Dick, W., Carey, L., & Carey, J. O. (2015). *The systematic design of instruction* (8th ed.). Pearson.
Eka Setiawati, N. L., Era Marsakawati, N. P., & Ramendra, D. P. (2024). Authentic assessment in action: English teachers' perception and practices for assessing students' productive skills at senior high school. *Indonesian Journal of Educational Development (IJED)*, *5*(3), 366–375. https://doi.org/10.59672/ijed.v5i3.4219
Habibi, A., Mukminin, A., & Sofyan, S. (2024). Access to digital technology in vocational schools. *Education and Information Technologies, 29*, 4197–4222. https://doi.org/10.1007/s10639-023-12006-x
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
Keller, J. M. (2010). *Motivational design for learning and performance*. Springer. https://doi.org/10.1007/978-1-4419-1250-3
Koo, T. K., & Li, M. Y. (2021). Guidelines for selecting and reporting intraclass correlation coefficients. *Journal of Chiropractic Medicine, 20*(2), 155–163. https://doi.org/10.1016/j.jcm.2021.02.012
Leppink, J. (2020). Cognitive load theory in educational research. *Educational Psychology Review, 32*, 1–15. https://doi.org/10.1007/s10648-019-09500-0
Martin, F., Sun, T., & Westine, C. (2020). Systematic review of online teaching. *Computers & Education, 159*, 104009. https://doi.org/10.1016/j.compedu.2020.104009
Mayer, R. E. (2014). Cognitive theory of multimultimedia learning. Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

Megawaty, M., Sulistiyo, U., Wachyuni, S., & Sofyan. (2025). Development of M.E.G.A: A self-assessment expert system to improve English correspondence skills for job-seeking university students. *Indonesian Journal of Educational Development (IJED)*, *6*(2), 393–407. https://doi.org/10.59672/ijed.v6i2.5099

Sofyan, S., Habibi, A., Sofwan, M., et al. (2023). TPACK–UoTI validation. *Humanities and Social Sciences Communications, 10*, 55. https://doi.org/10.1057/s41599-023-01533-0

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Taber, K. S. (2023). Improving the use of cronbach's alpha. *Research in Science Education, 53*, 1021–1035. https://doi.org/10.1007/s11165-022-10052-4

Teo, T., & Zhou, M. (2021). Explaining intention to use technology. *Journal of Computing in Higher Education, 33*, 301–325. https://doi.org/10.1007/s12528-020-09259-1

Widana, I. W., Sopandi, A. T., Suwardika, I. G. (2021). Development of an authentic assessment model in mathematics learning: A science, technology, engineering, and mathematics (STEM) approach. *Indonesian Research Journal in Education, 5*(1), 192-209. https://doi.org/10.22437/irje.v5i1.12992

Widana, I. W. & Ratnaya, I. G. (2021). Relationship between divergent thinking and digital literacy on teacher ability to develop HOTS assessment. *Journal of Educational Research and Evaluation, 5*(4), 516-524. https://doi.org/10.23887/jere.v5i4.35128

Zamanzadeh, V., et al. (2020). Design and implementation of a content validity study: Development of an instrument. *Nursing Practice Today, 7*(4), 280–289. https://doi.org/10.18502/npt.v7i4.4031