

Transformation of Meme-Based Learning Assessment: Multimodal Assessment of Generation Z Student Artifacts

Fina Rifana¹, Rizki Hikmawan^{*2}

^{1,2} Department of Information System and Technology Education, Universitas Pendidikan Indonesia

Article Info	ABSTRACT
Article history: Received June 12, 2026 Revised June 17, 2026 Accepted June 19, 2026 Corresponding Author: Rizki Hikmawan, Universitas Pendidikan Indonesia, Bandung, Jawa Barat, Indonesia. Email: hikmariz@upi.edu	Digital transformation has reshaped how Generation Z students communicate knowledge, prompting a re-examination of assessment practices in higher education. Assessment must evolve beyond traditional text-based tasks to accommodate multimodal literacies, yet empirical investigation of rubric-based dimensions in student-created multimodal artifacts remains underexplored. This study analyzes the quality of student meme artifacts as multimodal assessments and examines statistical relationships among assessment dimensions. A quantitative descriptive-correlational design was employed to analyze 155 student-generated meme artifacts. The analysis used a researcher-developed six-dimensional rubric to assess the data. Data were analyzed using descriptive statistics and Pearson correlation via IBM SPSS. Results indicate that overall meme quality falls predominantly in the Very High category across five of six dimensions, with Creativity and Originality as the sole dimension in the High category, with Ethics and Audience Appropriateness ($M = 3.99$) and Scientific Accuracy ($M = 3.81$) achieving the highest scores, while Creativity and Originality ($M = 3.20$) showed the greatest variation. Correlation analysis reveals significant positive relationships among most dimensions ($p < 0.01$). Multimodal Orchestration emerged as the central dimension, showing the strongest linkage with Creativity and Originality ($r = 0.643$) and Message Clarity ($r = 0.420$). These findings support the validity of meme-based assessment and highlight multimodal orchestration as a key factor in artifact quality. Keywords: multimodal assessment; meme; learning artifact; generation Z; higher education

1. INTRODUCTION

Digital transformation has fundamentally reshaped how Generation Z students construct and communicate knowledge, prompting a critical re-examination of assessment practices in higher education. The assessment as learning paradigm repositions assessment as an active mechanism through which students reflect on, monitor, and construct their understanding (Holmes et al., 2021; Nieminen, 2024). Within this framework, assessments are designed to elicit authentic demonstrations of students' thinking processes through contextual tasks (Wong & Liem, 2022). The artifacts produced through such tasks serve as representational evidence of conceptual understanding and learning experiences (Tang & Hew, 2022). As digital learning environments evolve, forms of student representation are increasingly multimodal: students integrate visual, symbolic, and other semiotic elements alongside linear text to convey meaning (Dahlström, 2022; G. Kress, 2009; Wanselin et al., 2022). Multimodal assessment facilitates the expression of understanding through the orchestration of multiple, complementary semiotic modes (Sidekierskienė & Damaševičius, 2025; Subbiramaniyan et al., 2022; Tidy et al., 2024). This approach is particularly well-aligned with the characteristics of Generation Z students, who have grown up in visual-based digital cultures characterized by rapid information flow and participatory communication practices on social media (Lamba & Jain, 2025). One multimodal artifact dominant in the

communicative repertoire of this generation is the internet meme, which combines brief text, visuals, and contextual humor to deliver culturally resonant messages in a concise yet meaningful way (Knobel & Lankshear, 2018; Oswald et al., 2023; Tyrer, 2021).

In the context of learning, memes have been utilized as a medium to enhance engagement, reflection, and connection between academic material and students' social experiences (Subbiramaniyan et al., 2022; Tidy et al., 2024). Several studies indicate that meme-based tasks can help students accurately reflect the quality of meaning representation and conceptual understanding in a personal and contextual manner (Sidekerskienė & Damaševičius, 2025). However, the use of memes as formal learning artifacts demands a clear evaluation framework so that assessment transcends the entertainment dimension and accurately reflects students' academic understanding. Although previous research has developed rubrics for assessing memes as digital multimodal compositions, these studies generally focus on pedagogical and descriptive aspects. The study by Ryu et al. (2022) for example, formulated a genre-based rubric to assess memes in second language learning contexts, emphasizing multimodality, culture, language, and genre function. While that rubric illuminated student performance in meme creation, it was not directed towards empirical testing of the assessment construct structure or the statistical relationships among dimensions. Consequently, assessment dimensions in meme-based rubrics are still treated as independent evaluative categories, not yet empirically tested as interrelated variables that collectively shape the overall quality of multimodal artifacts. A review of meme-based assessment studies (2018–2025) reveals that no study has yet applied correlation analysis to examine dimensional interdependencies within a rubric-based meme assessment framework in formal higher education settings.

Despite the growing body of research on meme-based learning, existing studies predominantly emphasize memes as instructional or engagement tools rather than as formal assessment instruments (Sidekerskienė & Damaševičius, 2025; Subbiramaniyan et al., 2022; Tidy et al., 2024). Previous research has largely focused on students' motivation and creativity in meme production, while empirical investigations examining memes as structured, rubric-based assessment artifacts with tested dimensional interrelationships remain limited. Unlike these prior studies, this research uniquely positions meme quality as a multidimensional empirical construct within formal higher education assessment in Indonesia. Specifically, it is among the first studies in the Indonesian higher education context to empirically examine the interrelational structure of six rubric-based assessment dimensions using Pearson correlation analysis, thereby offering an evidence-based foundation for meme assessment validity. Rather than merely replicating existing frameworks, this study bridges creative multimodal expression with the demands of accountable, standardized academic evaluation (Brunfaut & Kormos, 2025; Insulander et al., 2022; Wang et al., 2023). Therefore, this study aims to: (1) describe the quality of student meme artifacts across six rubric-based assessment dimensions, (2) examine the correlational structure among assessment dimensions, and (3) identify the most central assessment dimension within the multimodal artifact quality construct.

2. RESEARCH METHOD

This study employs a quantitative approach with a descriptive-correlational design. A descriptive-correlational design is appropriate for characterizing the quality distribution of student artifacts and testing hypothesized relationships among assessment dimensions without experimental manipulation (Creswell & Creswell, 2018). Each assessment dimension is positioned as a quantitative variable analyzed empirically, enabling both a descriptive account of meme quality and an examination of dimensional interdependencies.

2.1. Sampling and Population

The population of this study consisted of all undergraduate students enrolled in a technology-based education course at a state university in Indonesia, totaling 163 students. The course integrates foundational educational psychology concepts, including motivation theory, into technology-based learning activities. Students were assigned to create memes representing these concepts as part of the formal assessment, making the meme topics directly relevant to the course content. The research was conducted from October to November

2025. The sampling technique used was total sampling, as all students were required to complete the meme creation task as part of the formal course assessment. From the total population, 5 students did not submit the task, and 3 meme artifacts could not be analyzed due to file access issues during the data collection and compilation process, thus not meeting the technical criteria for complete analysis. Therefore, 155 meme artifacts were included as the final research dataset (attrition rate: 4.9%).

Although the meme creation task was mandatory as part of the course assessment, the use of task artifacts for research purposes was voluntary and did not affect students' academic grades. All data were anonymized prior to analysis, and no identifying information was retained or published. This study was conducted in accordance with the institutional research ethics guidelines of the affiliated university. Student consent for the use of their artifacts in research was obtained through an informed consent procedure at the beginning of the course.

2.2. Data Collection and Analysis

Data collection was conducted through a structured assignment procedure. Students were asked to create one educational meme that articulated course material. Students were permitted to use any digital tool for meme creation, including online meme generators (e.g., Imgflip, Canva) or graphic design applications. No restrictions were imposed on text length; however, students were instructed to ensure that all textual and visual elements were directly related to the course material. These design parameters were communicated through a written assignment brief distributed at the beginning of the task. Completed memes were then uploaded to the Instagram Story feature on students' personal accounts, a platform selected for its familiarity among Generation Z learners and its support for visual-based digital communication. After the upload period ended, students submitted documentation of the memes in the form of screenshots or links into a provided Google Drive folder. All accessible meme artifacts were then compiled by the researcher for analysis.

The research instrument was an analytical rubric for assessing memes, developed through a systematic adaptation process from the digital multimodal composition assessment framework proposed by Ryu et al. (2022). Prior to data collection, the adapted rubric underwent a content review process to ensure its relevance to the Indonesian higher education context; feedback received was incorporated into the final instrument.

The six dimensions of the rubric assess distinct yet complementary aspects of meme quality. It is important to distinguish between two dimensions that may appear similar: Content Relevance refers to the degree to which the meme's topic aligns with the learning objectives and subject matter of the course, whereas Scientific Accuracy refers to the factual and conceptual correctness of the information presented, namely whether the academic claims embedded in the meme are free from errors or misconceptions. A meme may be topically relevant but still contain conceptual inaccuracies, and vice versa. The complete rubric with scoring descriptors and weighting is presented in Table 1.

Table 1. Adapted Meme Assessment Rubric Developed by the Researcher

Criteria (Brief Description)	Score 4 (Excellent)	Score 3 (Good)	Score 2 (Fair)	Score 1 (Poor)	Weight (%)
1. Content Relevance (Alignment with learning objectives and subject matter)	Accurately and thoroughly represents the core concepts of the learning material; no misconceptions are present.	Relevant and generally accurate, with minor simplifications.	Related to the topic but superficial or contains minor inaccuracies.	Irrelevant or misleading with respect to the learning material.	25%
2. Message Clarity (Comprehensibility for the target audience)	Message is direct, clear, and easily understood by the target audience without additional explanation.	Message is generally clear but requires minimal contextual explanation.	Message is somewhat ambiguous and requires significant explanation.	Message is confusing or not successfully conveyed.	20%
3. Multimodal Orchestration (Integration of text, images, and layout)	Text, images, and layout are harmoniously integrated (concurrent, effective; a small	Integration is generally effective; a small	Multiple modes are present, but orchestration is	Modes are poorly integrated or	20%

Criteria (Brief Description)	Score 4 (Excellent)	Score 3 (Good)	Score 2 (Fair)	Score 1 (Poor)	Weight (%)
images, and visual design)	complementary, or purposefully divergent); modal choices strongly support meaning-making.	number of modes are not fully optimized.	weak; some elements are irrelevant or distracting.	interfere with comprehension.	
4. Scientific / Conceptual Accuracy (Factual and conceptual correctness)	No conceptual or factual errors; arguments and representations are academically accurate.	Minor conceptual issues that do not significantly affect meaning.	Errors partially affect conceptual understanding.	Serious conceptual errors that mislead or confuse the audience.	15%
5. Creativity & Originality (Innovative ideas, humor, and meaningful remixing)	Highly original; humor is relevant and demonstrates creative thinking that enhances conceptual understanding.	Creative and engaging, although using common meme templates.	Limited creativity; relies heavily on clichéd or overused elements.	Lacks originality; simple copy-paste without meaningful adaptation.	10%
6. Ethics & Audience Appropriateness (Cultural sensitivity and ethical considerations)	Fully appropriate and culturally sensitive; adheres to ethical standards and is non-offensive.	Generally appropriate with minimal risk of misinterpretation.	Potential cultural or ethical issues; requires greater sensitivity.	Contains offensive, discriminatory, or inappropriate content.	10%
Total					100%

The adapted rubric consists of six assessment dimensions: (1) content relevance, (2) message clarity, (3) multimodal orchestration, (4) scientific accuracy, (5) creativity and originality, and (6) ethics and audience appropriateness. Each dimension is assessed using a 1–4 scale with scoring weights adjusted according to learning priorities. All meme artifacts were assessed by one rater, the researcher, using the standardized analytical rubric. To minimize potential subjectivity inherent in single-rater assessment of creative artifacts, a structured calibration procedure was implemented prior to scoring. Specifically, the researcher independently scored 10 anchor memes, selected to represent the full range of quality across all six dimensions, and systematically cross-checked each score against the behavioral descriptors in the rubric to ensure consistent and criterion-referenced application of scoring standards. Any score that could not be unambiguously matched to a rubric descriptor was reviewed and re-scored after re-reading the relevant criterion. This procedure was intended to reduce rater drift and promote internal consistency throughout the scoring process.

Quantitative data analysis was conducted using IBM SPSS Statistics (version 25). The analysis techniques included: (1) descriptive statistics to obtain mean values, standard deviations, and minimum and maximum values for each assessment dimension; (2) frequency distribution to map score distributions across each dimension; and (3) Pearson product-moment correlation analysis to examine the relationships among the six assessment dimensions, identifying patterns of interrelation in shaping meme artifact quality. Correlation magnitudes were interpreted using Cohen's (1988) benchmarks: $r \geq 0.50$ as large, $r = 0.30$ – 0.49 as moderate, and $r < 0.30$ as small. Mean scores were categorized as follows: 3.50–4.00 = Very High; 2.50–3.49 = High; 1.50–2.49 = Moderate; 1.00–1.49 = Low.

3. RESULT AND ANALYSIS

3.1. RESULT

3.1.1. Quality of Student Meme Artifacts

Table 2. Descriptive Statistics of Students' Meme Artifact Quality (N = 155)

	N	Minimum	Maximum	Mean	Std. Deviation
Content Relevance	155	1	4	3.72	.541
Message Clarity	155	1	4	3.72	.564
Multimodal Orchestration	155	1	4	3.53	.705
Scientific Accuracy	155	1	4	3.81	.532
Creativity & Originality	155	1	4	3.20	.724
Ethics & Audience Appropriateness	155	3	4	3.99	.080
Valid N (listwise)	155				

Note. Scores were rated on a scale from 1 (Poor) to 4 (Excellent).

Descriptive statistical analysis was performed on 155 student meme artifacts across six assessment dimensions. A summary is presented in Table 2. Based on the mean score conversion criteria applied in this study (3.50–4.00 = Very High; 2.50–3.49 = High; 1.50–2.49 = Moderate; 1.00–1.49 = Low), four of six dimensions fall within the Very High category, indicating that students were generally capable of producing academically appropriate and multimodally coherent meme artifacts.

The Ethics and Audience Appropriateness dimension obtained the highest mean score ($M = 3.99$; $SD = 0.08$), followed by Scientific Accuracy ($M = 3.81$; $SD = 0.53$). Both dimensions fall within the Very High category. The near-ceiling score on Ethics and Audience Appropriateness (with 154 of 155 artifacts receiving the maximum score) suggests a ceiling effect, likely attributable to students' familiarity with academic communication norms in digital spaces, consistent with the social media literacy characteristic of Generation Z learners (Lamba & Jain, 2025). The high Scientific Accuracy score affirms that students were able to represent course concepts correctly even within a popular and informal medium.

The Content Relevance and Message Clarity dimensions both recorded a mean score of $M = 3.72$ (Very High category), indicating that the memes created by students were generally relevant to course material and capable of conveying messages clearly to the target audience. Multimodal Orchestration showed a mean score of $M = 3.53$ (Very High category), reflecting that the integration of visual and verbal elements was performed well, though not uniformly optimized across all artifacts. This variation is evidenced by a comparatively higher standard deviation ($SD = 0.705$) relative to other high-performing dimensions.

The dimension with the lowest mean score was Creativity and Originality ($M = 3.20$; $SD = 0.72$), which falls in the High category based on the mean score conversion criteria. This result indicates considerable variation in students' creative expression, as confirmed by the highest standard deviation among all dimensions. While most students demonstrated adequate creativity, a notable proportion relied on familiar meme templates, namely widely circulated pop-culture stock images such as characters from SpongeBob SquarePants or the 'Two Buttons' format, in which students substituted only the text captions without meaningfully altering the visual composition or creating new imagery from scratch, rather than developing highly original multimodal compositions, suggesting that creativity in formal meme-based assessment requires more targeted pedagogical scaffolding.

Table 3. Frequency Distribution of Scores Across Assessment Dimensions

Dimension	Score 1	Score 2	Score 3	Score 4
Content Relevance	1	4	32	118
Message Clarity	2	3	31	119
Multimodal Orchestration	5	4	50	96
Scientific Accuracy	2	4	15	134
Creativity & Originality	6	10	86	53
Ethics & Audience Appropriateness	0	0	1	154

The frequency distribution of scores for each dimension is presented in Table 3. The distribution results confirm the ceiling effect on the Ethics and Audience Appropriateness dimension, where 154 of 155 artifacts (99.4%) received the maximum score of 4. In contrast, the Creativity and Originality dimension showed a more balanced distribution across score categories 3 and 4 (86 and 53 artifacts, respectively), indicating meaningful variation in the creative quality of student meme artifacts. The Multimodal Orchestration dimension similarly exhibited wider score dispersion than other dimensions, with 50 artifacts scoring 3 and 96 scoring 4.

To provide a clearer illustration of the scoring differences across assessment dimensions, two representative meme artifacts are presented below. Figure 1 presents an example of a meme artifact with limited creativity and originality, characterized by reliance on a widely-used meme template with minimal visual adaptation to course content. Figure 2, in contrast, presents a meme artifact that received the highest scores across all six dimensions, demonstrating harmonious integration of text and visual elements, conceptually accurate content, and a creative approach that goes beyond commonly used templates.



Figure 1. Example of a Low-Scoring Meme Artifact

Figure 1. This meme features two identical-looking characters with the captions 'aman?' and 'aman!' (Indonesian for 'safe?' and 'safe!'), with no discernible connection to course material. The minimal and context-free text fails to convey any academic concept, and the visual elements are not purposefully integrated to support meaning-making, resulting in low scores on Content Relevance, Message Clarity, and Multimodal Orchestration.



Figure 2. Example of a High-Scoring Meme Artifact (Score 4 across all dimensions)

Figure 2. Although this meme employs a recognizable pop-culture image (SpongeBob SquarePants characters), it distinguishes itself from low-scoring template-dependent memes through the purposeful and meaningful restructuring of the visual layout. Each character is mapped to a distinct stage of Self-Determination Theory, from Amotivation through Integrated Regulation, creating a visually coherent and conceptually accurate representation of the motivational continuum. It is the intentional adaptation of the template to serve an academic purpose, rather than mere substitution of text captions, that warranted high scores across all six dimensions including Multimodal Orchestration, Message Clarity, and Creativity and Originality.

3.1.2. Relationships Among Assessment Dimensions

Table 4. Pearson Correlation Matrix Among Meme Quality Assessment Dimensions

	Content Relevance	Message Clarity	Multimodal Orchestration	Scientific Accuracy	Creativity & Originality	Ethics & Audience Appropriateness
Content Relevance <i>Pearson r</i>	1	.639**	.319**	.608**	.358**	-.041
<i>Sig. (2-tailed)</i>		.000	.000	.000	.000	.609
<i>N</i>	155	155	155	155	155	155
Message Clarity <i>Pearson r</i>	.639**	1	.420**	.626**	.327**	-.040
<i>Sig. (2-tailed)</i>	.000		.000	.000	.000	.624
<i>N</i>	155	155	155	155	155	155
Multimodal Orchestration <i>Pearson r</i>	.319**	.420**	1	.283**	.643**	.061
<i>Sig. (2-tailed)</i>	.000	.000		.000	.000	.454
<i>N</i>	155	155	155	155	155	155
Scientific Accuracy <i>Pearson r</i>	.608**	.626**	.283**	1	.233**	.124
<i>Sig. (2-tailed)</i>	.000	.000	.000		.004	.126
<i>N</i>	155	155	155	155	155	155
Creativity & Originality <i>Pearson r</i>	.358**	.327**	.643**	.233**	1	-.089
<i>Sig. (2-tailed)</i>	.000	.000	.000	.004		.269
<i>N</i>	155	155	155	155	155	155
Ethics & Audience Appropriateness <i>Pearson r</i>	-.041	-.040	.061	.124	-.089	1
<i>Sig. (2-tailed)</i>	.609	.624	.454	.126	.269	
<i>N</i>	155	155	155	155	155	155

** . Correlation is significant at the 0.01 level (2-tailed).

To address the second research question, Pearson product-moment correlation analysis was conducted among the six assessment dimensions. The complete correlation matrix is presented in Table 4. Following Cohen's (1988) benchmarks for correlation magnitude ($r \geq 0.50$ = large; $r = 0.30-0.49$ = moderate; $r < 0.30$ = small), the analysis reveals positive and significant relationships among most assessment dimensions ($p < 0.01$), with the exception of the Ethics and Audience Appropriateness dimension.

Large-effect correlations were found between Content Relevance and Message Clarity ($r = 0.639$, $p < 0.01$), and between Message Clarity and Scientific Accuracy ($r = 0.626$, $p < 0.01$). These findings indicate that

memes artifacts with strong alignment to course material also tend to communicate messages more clearly and with greater conceptual accuracy—suggesting that content-related dimensions form a coherent quality cluster in student meme production. Additionally, the correlation between Content Relevance and Scientific Accuracy ($r = 0.608$, $p < 0.01$) further reinforces this pattern, indicating that academic quality in meme artifacts is characterized by a mutually reinforcing relationship among content, clarity, and accuracy.

Furthermore, Multimodal Orchestration showed a large-effect correlation with Creativity and Originality ($r = 0.643$, $p < 0.01$), and a moderate-effect correlation with Message Clarity ($r = 0.420$, $p < 0.01$). This pattern indicates that Multimodal Orchestration is closely associated with both creative expression and communicative effectiveness in meme artifacts. The moderate correlation between Multimodal Orchestration and Content Relevance ($r = 0.319$, $p < 0.01$) suggests that while this dimension supports content representation, it operates relatively independently from content-driven quality.

Conversely, the Ethics and Audience Appropriateness dimension did not statistically co-vary with any other assessment dimension (all $p > 0.05$). This finding is consistent with its ceiling effect: when a dimension has near-zero variance ($SD = 0.08$), it lacks the distributional spread necessary to produce meaningful correlations with other variables. This result suggests that Ethics and Audience Appropriateness functions as a normative prerequisite in meme-based assessment—a baseline condition that is uniformly met—rather than a dimension that interacts with variations in multimodal quality or artifact creativity. Future studies may consider whether this dimension is better conceptualized as a gatekeeping criterion rather than a continuous quality indicator.

3.1.3. Most Central Assessment Dimension

Based on the correlation matrix in Table 4, the Multimodal Orchestration dimension emerges as the most central dimension in the meme quality construct. It demonstrates significant correlations with four of the five other dimensions: a large-effect correlation with Creativity and Originality ($r = 0.643$), a moderate-effect correlation with Message Clarity ($r = 0.420$), and small-effect but statistically significant correlations with Scientific Accuracy ($r = 0.283$) and Content Relevance ($r = 0.319$). This pattern of consistent linkage positions Multimodal Orchestration as a bridging construct connecting the creative-expressive cluster (Creativity and Originality) and the content-communicative cluster (Content Relevance, Message Clarity, Scientific Accuracy).

This finding indicates that Multimodal Orchestration occupies a central and integrative position in shaping student meme quality as a multimodal assessment artifact. It is important to note that, as a correlational study, these findings describe patterns of co-variation among dimensions rather than causal or directional relationships. The centrality of Multimodal Orchestration should be interpreted as evidence of its structural importance within the meme quality construct, not as evidence that it causally produces quality in other dimensions.

3.2. ANALYSIS

The analysis of this study indicates that students were generally able to meet both normative and academic demands in creating memes as learning artifacts. The high scores on Ethics and Audience Appropriateness ($M = 3.99$) demonstrate strong sensitivity to audience context and academic communication norms in digital spaces. This aligns with the documented characteristics of Generation Z learners, who have developed sophisticated social media literacy through sustained participation in participatory digital cultures (Lamba & Jain, 2025; Oswald et al., 2023). However, the near-ceiling performance on this dimension—evidenced by an exceptionally low standard deviation ($SD = 0.08$)—raises a methodological consideration for future assessment design: when a rubric dimension produces near-zero variance, its discriminatory power as an assessment indicator is limited. Future researchers may consider repositioning Ethics and Audience Appropriateness as a prerequisite screening criterion rather than a weighted scoring dimension.

The high Scientific Accuracy scores ($M = 3.81$) reinforce the finding that the use of memes as a formal assessment medium does not necessarily compromise conceptual quality. Memes can function as vehicles for

articulating accurate academic understanding, as also evidenced in prior research on meme-based tasks in multimodal learning contexts (Ryu et al., 2022; Tidy et al., 2024). This finding challenges the concern among educators that popular or informal formats may lead students to oversimplify disciplinary content. Consistent with Sidekerskienė & Damaševičius (2025), who found that meme creation in STEM subjects supported accurate conceptual communication, the present results demonstrate that with appropriate rubric-based guidance, students can maintain academic rigor within informal multimodal genres. Unlike Subbiramaniyan et al. (2022) who focused primarily on engagement outcomes in physiology education, the present study demonstrates that meme quality can be measured with empirical precision across multiple academic dimensions simultaneously.

However, the comparatively lower score on Creativity and Originality ($M = 3.20$; $SD = 0.72$) indicates that a substantial proportion of students tended to reproduce commonly used meme templates rather than developing more innovative multimodal expressions. This is consistent with the broader literature on digital creativity in educational contexts, which notes that learners often default to familiar genre conventions when engaging in open-ended multimodal tasks (Dahlström, 2022; Knobel & Lankshear, 2018). The higher standard deviation on this dimension compared to others also indicates meaningful individual differences in creative capacity among students. These results suggest that creativity in formal meme-based assessment requires more explicit pedagogical support, including creative exemplars, structured scaffolding, and sufficient task exploration time prior to formal submission.

The large-effect correlation between Multimodal Orchestration and Creativity and Originality ($r = 0.643$) reinforces the view that creative expression in multimodal artifacts is deeply intertwined with Multimodal Orchestration, defined as the ability to integrate semiotic modes cohesively and purposefully. This finding aligns with Kress & Leeuwen (2001) social semiotic framework, which posits that meaning in multimodal texts is not made through any single mode in isolation, but through the orchestrated interplay of multiple semiotic resources—including visual, gestural, and linguistic elements. Similarly, (G. Kress, 2009) argues that design—the purposeful selection and combination of modes—is central to creative meaning-making in digital communication. The present data provide empirical support for these theoretical positions within a formal assessment context: students who demonstrated greater creative and original thinking also tended to show stronger orchestration of visual and textual elements. This pattern is consistent with findings by Ryu et al. (2022) who identified multimodal integration as the most distinguishing feature of high-quality student memes in L2 learning contexts.

The absence of significant correlations between Ethics and Audience Appropriateness and all other dimensions can be interpreted through the lens of ceiling effects and restricted range attenuation. When a variable has near-zero variance—as observed here with $SD = 0.08$ —it is statistically incapable of producing meaningful correlations regardless of its theoretical relationship with other constructs. The near-perfect Ethics scores ($M = 3.99$, $SD = 0.08$) suggest a ceiling effect that may be attributed to students' social media literacy and awareness of academic norms (Lamba & Jain, 2025; Oswald et al., 2023). This does not imply that ethical awareness is unimportant in meme-based assessment; rather, it suggests that this cohort uniformly met ethical standards. Future studies could examine ethics as a prerequisite variable rather than a predictor of artifact quality, particularly in contexts involving politically contentious or culturally sensitive topics where greater variance may be expected.

4. CONCLUSION

This study provides empirical evidence that memes can function as valid multimodal assessment artifacts capable of reflecting students' academic understanding with measurable consistency across multiple assessment dimensions. Based on the analysis of 155 student meme artifacts using six-dimensional meme assessment rubric, the overall quality of memes falls in the Very High to High category, particularly in Ethics and Audience Appropriateness ($M = 3.99$) and Scientific Accuracy ($M = 3.81$). These findings demonstrate that students are able to maintain academic norms and conceptual accuracy even within a popular and informal medium, affirming the potential of meme-based assessment as a formal evaluation instrument in higher education.

Correlation analysis indicates that meme artifact quality is shaped not by isolated assessment dimensions but by an interconnected network of multimodal competencies. Among these, Multimodal Orchestration emerged as the most structurally central dimension, showing strong relationships with Creativity and Originality as well as Message Clarity. These findings highlight that the effective integration of multiple semiotic modes plays a critical role in determining the overall quality of student-generated multimodal artifacts.

Theoretically, this study contributes to the multimodal assessment literature by providing empirical, correlation-based evidence of dimensional interdependencies within a meme assessment rubric, thereby supporting the construct validity of meme-based evaluation in formal educational settings. Practically, educators are encouraged to: (1) incorporate multimodal orchestration training in pre-task scaffolding; (2) use the six-dimensional rubric as a formative assessment guide throughout the creation process; and (3) provide students with creative exemplars demonstrating high creativity and originality to address the dimension showing greatest variation, thereby expanding students' multimodal repertoire beyond template-dependent production.

This study has several limitations that should be acknowledged. First, all meme artifacts were assessed by a single rater; while a calibration procedure was implemented, formal inter-rater reliability testing was not conducted, and this remains a significant methodological limitation to be addressed in future research. Second, data were collected from a single course at one institution, limiting the generalizability of findings across disciplines and educational contexts. Third, the cross-sectional correlational design precludes causal inferences about the relationships among assessment dimensions. Fourth, the study did not control for students' prior experience in meme creation, digital design, or content creation more broadly prior to enrolling in the course. Individual differences in pre-existing digital literacy and design skills may have substantially contributed to the high variance observed in Creativity and Originality scores ($SD = 0.72$), and future studies should consider including a pre-task survey to capture such background variables. Subsequent research should address these limitations, involve multiple raters, and explore meme-based assessment in diverse disciplinary contexts to strengthen the external validity of findings. Future studies are also encouraged to replicate this assessment framework across other subject areas, such as STEM or social sciences, and to employ mixed-methods approaches that incorporate student interviews or think-aloud protocols to better understand the cognitive and creative processes underlying meme artifact production.

ACKNOWLEDGEMENTS

The authors would like to express sincere gratitude to all students who participated in this study and willingly contributed their meme artifacts as research data. This research received no external funding and was conducted independently as part of the authors' academic work at Universitas Pendidikan Indonesia.

DECLARATION OF ARTIFICIAL INTELLIGENCE USE IN THE WRITING PROCESS

The authors affirm that this manuscript is the product of their own intellectual work, including the research design, data collection, analysis, interpretation of findings, and academic argumentation. Artificial intelligence (AI) tools, Claude (Anthropic, Claude Sonnet), were used in a limited capacity to assist with language editing and grammatical refinement of selected passages. All substantive content, scholarly judgment, and conclusions presented in this manuscript were independently developed and remain the sole intellectual contribution of the authors, who take full responsibility for the accuracy, integrity, and originality of this work.

REFERENCES

- Brunfaut, T., & Kormos, J. (2025). Assessing Multimodal Viewing-to-Write Constructs: Task Design, Performance, Processing, and Rating. *Language Assessment Quarterly*, 22(4–5), 429–459. <https://doi.org/10.1080/15434303.2025.2596374>

- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (Fifth edition). SAGE.
- Dahlström, H. (2022). Students as digital multimodal text designers: A study of resources, affordances, and experiences. *British Journal of Educational Technology*, 53(2), 391–407. <https://doi.org/10.1111/bjet.13171>
- Holmes, A. G. D., Polman Tuin, M., & Turner, S. L. (2021). Competence and competency in higher education, simple terms yet with complex meanings: Theoretical and practical issues for university teachers and assessors implementing Competency-Based Education (CBE). *Educational Process International Journal*, 10(3). <https://doi.org/10.22521/edupij.2021.103.3>
- Insulander, E., Hernwall, P., Åkerfeldt, A., & Öhman, E. (2022). Assessment in Transformation: Teachers' Perceived Opportunities and Challenges in the Assessment of Multimodal Texts. *Educare*, (1), 132–153. <https://doi.org/10.24834/educare.2022.1.6>
- Knobel, M., & Lankshear, C. (2018). *Online memes, affinities and cultural production (2018 update to our 2007 chapter)* To appear as: Knobel, M. and Lankshear, C. (forthcoming). *Memes online, afinidades e produção cultural (2007 – 2018)*. In Chagas, Viktor (ed.). *Estudos sobre Memes: história, política e novas experiências de letramento*. 2019. <https://doi.org/10.13140/RG.2.2.34717.77280>
- Kress, G. (2009). *Multimodality* (0 ed.). Routledge. <https://doi.org/10.4324/9780203970034>
- Kress, G. R., & Leeuwen, T. van. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.
- Lamba, P. S., & Jain, N. (2025). Psycho-physiological impact of virtual non-verbal communication on Gen Z workforce: A study of memes. *Acta Psychologica*, 254, 104848. <https://doi.org/10.1016/j.actpsy.2025.104848>
- Nieminen, J. H. (2024). Assessment for Inclusion: Rethinking inclusive assessment in higher education. *Teaching in Higher Education*, 29(4), 841–859. <https://doi.org/10.1080/13562517.2021.2021395>
- Oswald, E. C., Esborg, L., & Pierroux, P. (2023). Memes, youth and memory institutions. *Information, Communication & Society*, 26(15), 2999–3016. <https://doi.org/10.1080/1369118X.2022.2131363>
- Ryu, J., Kim, Y. A., Eum, S., Park, S., Chun, S., & Yang, S. (2022). The assessment of memes as digital multimodal composition in L2 classrooms. *Journal of Second Language Writing*, 57, 100914. <https://doi.org/10.1016/j.jslw.2022.100914>
- Sidekerskienė, T., & Damaševičius, R. (2025). Pedagogical memes: A creative and effective tool for teaching STEM subjects. *International Journal of Mathematical Education in Science and Technology*, 56(7), 1230–1260. <https://doi.org/10.1080/0020739X.2024.2328818>
- Subbiramaniyan, V., Apte, C., & Ali Mohammed, C. (2022). A meme-based approach for enhancing student engagement and learning in renal physiology. *Advances in Physiology Education*, 46(1), 27–29. <https://doi.org/10.1152/advan.00092.2021>
- Tang, Y., & Hew, K. F. (2022). Effects of using mobile instant messaging on student behavioral, emotional, and cognitive engagement: A quasi-experimental study. *International Journal of Educational Technology in Higher Education*, 19(1), 3. <https://doi.org/10.1186/s41239-021-00306-6>

- Tidy, H., Bolton-King, R. S., Croxton, R., Mullen, C., Nichols-Drew, L., Carlysle-Davies, F., Moran, K. S., & Irving-Walton, J. (2024). Enhancing the student learning experience through memes. *Science & Justice*, 64(3), 280–288. <https://doi.org/10.1016/j.scijus.2024.03.004>
- Tyrer, C. (2021). The voice, text, and the visual as semiotic companions: An analysis of the materiality and meaning potential of multimodal screen feedback. *Education and Information Technologies*, 26(4), 4241–4260. <https://doi.org/10.1007/s10639-021-10455-w>
- Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023). Chinese EFL Teachers' Writing Assessment Feedback Literacy: A Scale Development and Validation Study. *Assessing Writing*, 56, 100726. <https://doi.org/10.1016/j.asw.2023.100726>
- Wanselin, H., Danielsson, K., & Wikman, S. (2022). Analysing Multimodal Texts in Science—A Social Semiotic Perspective. *Research in Science Education*, 52(3), 891–907. <https://doi.org/10.1007/s11165-021-10027-5>
- Wong, Z. Y., & Liem, G. A. D. (2022). Student Engagement: Current State of the Construct, Conceptual Refinement, and Future Research Directions. *Educational Psychology Review*, 34(1), 107–138. <https://doi.org/10.1007/s10648-021-09628-3>