





Multimodal reading comprehension competence in university students from social sciences and engineering

 **Guido Torres-Orihuela**¹

 **Miriam-Rosario Flores-Castro Linares**²

 **Ferly Elmer Urday Luna**³⁺

^{1,2,3} Universidad Católica de Santa María, Arequipa, Peru.

¹Email: htorres@ucsm.edu.pe

²Email: mflores@ucsm.edu.pe

³Email: furday@ucsm.edu.pe



(+ Corresponding author)

ABSTRACT

Article History

Received: 29 July 2025

Revised: 2 February 2026

Accepted: 20 February 2026

Published: 3 April 2026

Keywords

Academic performance
Disciplinary literacy
Multimodal reading comprehension
Multisemiotic competence
Semiotic modes
Text perception
University students.

This study investigates multimodal reading comprehension competence among university students in social sciences and engineering programs, focusing on the influence of cognitive, attitudinal, and academic variables. Adopting a quantitative, non-experimental, and correlational design, the research employed Poisson regression analysis to examine how different semiotic modes verbal, graphic, and combined affect comprehension outcomes. The study also considered prior experience with text formats, self-reported comfort levels, and academic progression as predictor variables. The sample consisted of 426 students from a private Peruvian university, stratified by discipline and semester level. Results show that students who received multimodal texts achieved significantly higher scores, suggesting that multimodal integration enhances comprehension through dual-channel processing. Additionally, positive perception of text comfort and previous exposure to verbal formats were positively associated with performance. Conversely, students with stronger mathematical or graphic backgrounds and those in advanced academic semesters performed less well, indicating potential gaps in multimodal literacy development across disciplines and stages of study. These findings underscore the importance of integrating multisemiotic literacy as a transversal academic skill, not only as a technical reading competence but as a critical component of epistemic participation in higher education. The study highlights the need for curricula to explicitly address multimodal comprehension strategies and to provide students with guided practice in interpreting and producing complex multimodal artifacts.

Contribution/Originality: This study contributes to the existing literature by examining multimodal reading comprehension among Peruvian university students. Using Poisson regression, it integrates cognitive, attitudinal, and disciplinary variables, demonstrating how modality, prior experiences, and comfort interact. It provides localized evidence and advocates for embedding multimodal literacy as a transversal competence in university curricula.

1. INTRODUCTION

Multimodality has become increasingly important in research related to education, and communication, in general, ends up being equally important because it acknowledges the complexity involved in the meaning-making of these modes of communication, now widely accepted in today's texts. The reading comprehension that takes place in universities may not only involve verbal language but also increasingly involves images, diagrams, sounds, non-linear text, interactive elements, etc. The current research being conducted has presented theoretical models and

classroom descriptions (For example, (Bateman, 2019; Kress & Van Leeuwen, 2006; Mayer, 2014; Parodi & Julio, 2015)), but there is a lack of empirical research that examines the ways in which university students in different disciplines engage with multimodal texts, whether in textbooks, notebooks, or online. Available research addressing multimodal literacy has largely focused on theoretical models surrounding multimodal texts or on early and secondary education, which means that higher education, especially in Latin America, remains unaccounted for. Additionally, there is limited previous research regarding experiential and attitudinal factors, such as prior exposures to multimodal resources or students' perceived comfort when dealing with multimodal texts across disciplines.

This research study aims to address some of these gaps by analyzing multimodal comprehension in undergraduate students in social sciences and engineering in Peru. The study intends to: (1) analyze the influence of text type (verbal, graphic, and combined) on reading comprehension performance; (2) examine the predictive value of experiential and attitudinal factors such as perceived comfort with certain types of text and prior exposures to different text types; and (3) reflect on what constitutes multimodal literacy in higher education. In doing so, this study seeks to provide empirical evidence to augment the theoretical and empirical discussions surrounding this knowledge and emphasizes that multimodal literacy must be embraced as a transversal competency in university education.

2. LITERATURE REVIEW

Multimodality has evolved to become an important conceptual lens for understanding the construction of meaning in contemporary texts. Meaning is no longer only instantiated through written and spoken language; it emerges when semiotic resources such as images, gestures, color, spatial design, and sound mediate with each other socially and culturally. Kress and Van Leeuwen (2001) and Kress and Van Leeuwen (2006) set the precedent for this lens by establishing that texts should not be treated or understood as linear or verbal texts but as multimodal texts containing various modes of communication serving various communicative functions. Following Kress and Van Leeuwen (2006); Bateman (2014) and Bateman (2019) further developed their work as scholars explored multimodal analysis as it related to reading or pedagogy beyond visual or verbal cues, and began to emphasize that "multimodal analysis is systematic and we need models that can replicate such complex relations." Even with Kress & Van Leeuwen and Bateman's contributions, the activity of knowing and understanding how to delineate the boundaries of semiotic modes and types of separable modes is still the focus of much academic debate and theorization (Pflaeging, Bateman, & Wildfeuer, 2021).

This theoretical approach has greatly influenced the field of education, particularly since students are often accessing materials in which graphics, diagrams, charts, symbols, digital resources, and words converge. As noted, reading is understood differently from the decoding of linear text; in higher education, it also constitutes an expectation that students can synthesize information across the semiotic channels. Mayer (2009) and Mayer (2014), the cognitive theory of multimedia learning supports this approach to learning and comprehension, where logical arguments (dual processing) are coupled with images (or visuals) and presented coherently to facilitate comprehension and meaningful knowledge transfer. At the same time, Mayer warns that texts that are incoherent or have poor visual designs can create cognitive overload for students, which is consistent with the issues described by Jewitt (2008); Jewitt (2009) and Bezemer and Kress (2016) whereby multimodal resources presented to students under poorly designed circumstances can lead to confusion and disorientation without explicit instruction to help students make connections across the modes. Teaching in multimodal environments highlights the opportunities and risks that students face and exposes the need to include features in their representations to help them craft their responses.

For the purposes of this discussion, comprehension is more than the retrieval of the explicit. Commonly agreed-upon definitions of comprehension, literal, inferential, critical, and reconstructive, take on different meanings when multimodal texts are considered (Barrett, 1976; Kintsch, 1998; McNamara & Magliano, 2009; Snow, 2002). Application of literal comprehension, for example, could possibly mean recognizing not only what has been written (i.e., words in the text), but also what is present in graphical form (i.e., visually represented) through visuals, text,

tables, or diagrams. Inferential comprehension can suggest cross-links of verbal and visual representations which coordinate to create an implicit meaning, such as knowing a cause-and-effect relationship is occurring based on the causal diagram or process chart. Critical comprehension could engender questioning the integrity of the interaction of the modes, such as concomitantly determining if the figure is reinforcing or contradictory to the written text, and how the displayed design elements may help shape the audience's comprehension of the text. Finally, reconstructive comprehension is more extensive than merely depiction since it leads to reconstructing the multimodal input text to generate a new output text (e.g., designing an infographic based on a scientific paper or multimodal summary of a lecture). These dimensions of comprehension indicate that the cognitive load is high in multimodal reading, as the skills associated are beyond the confines of traditional literacy.

There is evidence from empirical research to suggest that students often find these demands difficult. Eye-tracking studies conducted by Parodi and Julio (2017) demonstrate that while the university students they studied were adept at reading, they tended to fixate on a verbal text to the exclusion of visual elements, which is referred to as logocentrism, feeding into a logocentric reading process and suggesting that educational traditions emphasize the centrality of written language and limit potentially composite reading processes or the reception and action on the information via the multimodal semiotic resources. Although a number of studies have examined further aspects shaping the way multimodal texts are understood, for example, Gladic-Miralles and Cautín-Epifani (2018) found that students with a greater degree of experience or confidence in digital modes demonstrated more proficiency in integrating verbal with visual data, whereas students who were more mathematical or less confident performed better with regard to inter-modal connections. Most recently, De-la-Peña, Chaves-Yuste, and Luque-Rojas (2024) made the case that digital exposure was necessary but not sufficient for students to display a sophisticated degree of multimodal literacy, and that while they may click on hyperlinks and interact with media, students make the connection on a surface level without demonstrating critical comprehension. These findings all support the need to explicitly teach how to develop multimodal literacy potency, rather than consider an awareness of digital spaces as sufficient.

The implications for higher education are hard to overstate. For example, Cope and Kalantzis (2009) outline through their theory of multiliteracies that students must be positioned to function in environments differentiated by culture and technology. Similarly, Shanahan and Shanahan (2008) suggest that disciplinary literacy requires educators to be mindful of the ways in which it differs from general reading, and, importantly, the multimodal conventions of that disciplinary literacy. Additionally, Parodi (2015) has supported this position by investigating how academic text operates with semiotic configurations as contextually and discipline-dependent; therefore, students are required to make sense of modes as epistemically constrained. In their studies between 2016 and 2017, Parodi and Julio (2017) empirically consider the discursive challenges that their study subjects, students, displayed from corpus-based analysis of their discourse, as well as the eye-tracking data. Martínez Lirola and Ibañez Castejon (2015) and Pinheiro (2016) contend that higher education needs not only support students to interpret visual resources but also demonstrate how to produce multimodal artefacts such as posters, infographics, and interactive presentations, illustrating that they are more than academic competencies; they are professional ones, and students will be expected to perform each "mode" as part of their contribution to an increasingly multimodal workplace, whether academic or professional.

Multimodal reading cognition involves dual coding, cross-association, and intermodal inference processes. Multimodal reading also utilizes attention control over an extended period of time. Mayer's coherence principle and contiguity principle Mayer's (2014) still apply, but researchers are cautioned, like Serafini (2012) and Danielsson and Selander (2021) that multimodal literacy cannot be assumed to be implicit. Serafini's three-dimensional model of literacy involves the reader embracing all three levels of literacy, including decoding, analyzing interactions, and critical evaluation. Danielsson and Selander (2021) advocated for multisemiotic literacy across all educational stages, and an integrated approach will provide the framework to position students with greater confidence and contextual experience working with multimodal texts through higher education. Overall, the collective arguments have made it

clear that multimodal literacy, as described, is far from a peripheral skill and is arguably a 'basic' component of contemporary education.

In conclusion, the research suggests that multimodality changes the reading, learning, and knowledge production process in higher education. The problem is that most students come to university with inadequate preparation for working with multimodal texts, and the academic and professional requirements of the 21st century demand the ability to work.

At the same time, this is a blessing in disguise because by intentionally addressing multimodal literacy in the curriculum, educators can improve understanding, facilitate critical thinking, and prepare students to engage fully in disciplinary and civic communities. This study situates itself in this context, providing empirical evidence of how multimodal comprehension operates in the university context for students and what pedagogies might facilitate multimodal comprehension.

3. METHODOLOGY

3.1. Approach and Research Design

This study employed a quantitative, non-experimental, cross-sectional, and correlational research design. It explored the relationship between multimodal reading comprehension performance and a set of predictor variables, including text type, prior experience, perceived comfort, and academic semester. Statistical modeling techniques were applied to estimate the relative effect of each variable on the dependent outcome.

3.2. Participants

A total of 426 undergraduate students from Universidad Católica de Santa María (UCSM), a private university in southern Peru, participated in the study.

Participants were enrolled in Social Sciences and Engineering programs. A stratified probabilistic sampling strategy ensured balanced representation by academic semester and discipline. All participants provided informed consent under institutional ethical protocols.

3.3 Instruments

Three main instruments were used in data collection.

- Multimodal Reading Comprehension Test: Assessed students' ability to interpret texts with three modal configurations: verbal, graphic, and combined. The instrument comprised 13 multiple-choice items measuring literal, inferential, and critical comprehension based on a statistical text.
- Reading Experience and Preference Questionnaire: Captured participants' familiarity with multimodal texts, their frequency of exposure, and their preferred reading modality.
- Perceived Comfort Scale: Evaluated self-perceived ease in interpreting multimodal content using a four-point ordinal scale ranging from 'very comfortable' to 'not at all comfortable'.

The analysis focused on three text types: (1) exclusively verbal; (2) exclusively graphic; and (3) combined texts aligned with Mayer's (2014) principles of coherence and contiguity. The instruments were grounded in semiotic theory and designed to measure multimodal comprehension across cognitive and emotional domains.

3.4. Instrument Validation

The comprehension test was carefully constructed and validated before the main application. The verbal version presented short academic passages; the graphic version included tables and diagrams without accompanying text; and the combined version integrated verbal and visual resources according to Mayer's (2014) principles of coherence and contiguity. A pilot test with 30 students was carried out to check clarity, timing, and item difficulty, leading to minor adjustments. In addition, three specialists in linguistics and education reviewed the items to ensure content validity

and alignment with the three levels of comprehension. The decision to use statistical texts was intentional, as this type of material commonly integrates verbal and graphic information, offering a suitable basis for measuring multimodal reading skills.

3.5. Reliability Analysis

Internal consistency was evaluated using Cronbach's alpha (α) and McDonald's omega (ω).

- Semiotic Modes Perception Scale: Eight Likert-type items with $\alpha = 0.743$ and $\omega = 0.744$, considered acceptable for exploratory purposes.
- Multimodal Reading Comprehension Scale: Thirteen items with $\alpha = 0.765$ and $\omega = 0.811$, indicating strong reliability. Item analysis suggested minor alpha gains if items P10 and P13 were removed; however, they were retained for construct integrity.

3.6. Procedure

Instruments were administered in person under controlled conditions. Students completed the comprehension test first, followed by the questionnaire and comfort scale. Text type assignment was randomized:

- 148 received verbal texts.
- 123 received graphic texts.
- 155 received combined texts.

All responses were anonymized and processed in RStudio in accordance with ethical research standards.

3.7. Data

The data consisted of students' responses to the three instruments described above. These were collected from Social Sciences and Engineering undergraduates at UCSM, southern Peru. The dataset integrated comprehension scores, self-reported reading experience, and comfort ratings. All data were anonymized and stored securely, with access limited to the research team.

3.8. Statistical Analysis

A Poisson regression model was used to analyze the relationship between predictor variables and the total number of correct responses. Model selection followed a stepwise approach with validation through overdispersion checks (AER), zero-inflated comparison (PSCL), residual diagnostics (DHARMA), and influence metrics (Cook's index).

4. RESULTS

4.1. Descriptive Analysis

Multimodal reading comprehension scores varied according to the type of text received. The overall median was 9 points out of a maximum of 13. Figure 1 shows the distribution of scores by text type: verbal, graphic, or combined (Verbal + Graphic).

Students who received the combined text obtained the highest scores (Median = 10), followed by those who read exclusively graphic texts (Median = 9) and verbal texts (Median = 8). The combined group also showed less score dispersion, suggesting a positive effect on both performance level and consistency.

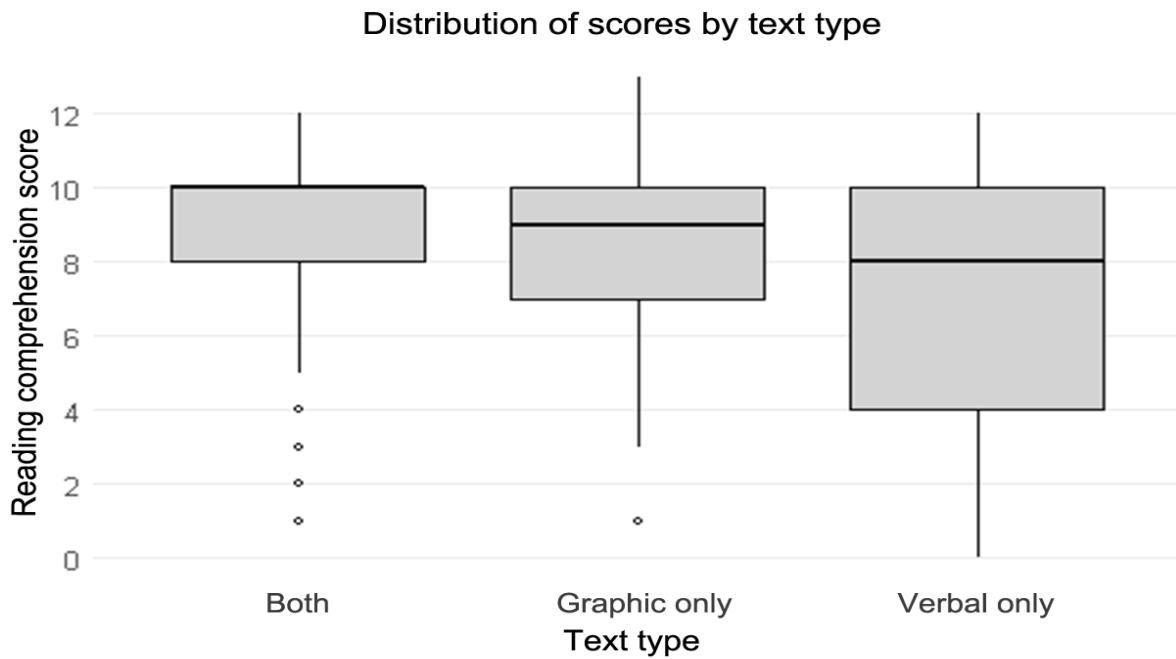


Figure 1. Score distribution by text type.

Note: The boxplot presents reading comprehension score distributions by text type: verbal, graphic, or combined (verbal + graphic).

A significant positive effect was observed for combined texts (verbal + graphic) and prior verbal experience. In contrast, mathematical experience, advanced academic semester, and perceived discomfort showed negative effects.

Regarding comfort perception, Figure 2 displays comprehension score distributions based on participants' self-reported comfort levels. Students who reported feeling "very comfortable" achieved higher scores than those who felt "slightly" or "not at all comfortable." However, an interesting pattern emerged: those who reported being 'not at all comfortable' showed the highest median scores and the least dispersion, suggesting a non-linear relationship between comfort perception and performance.

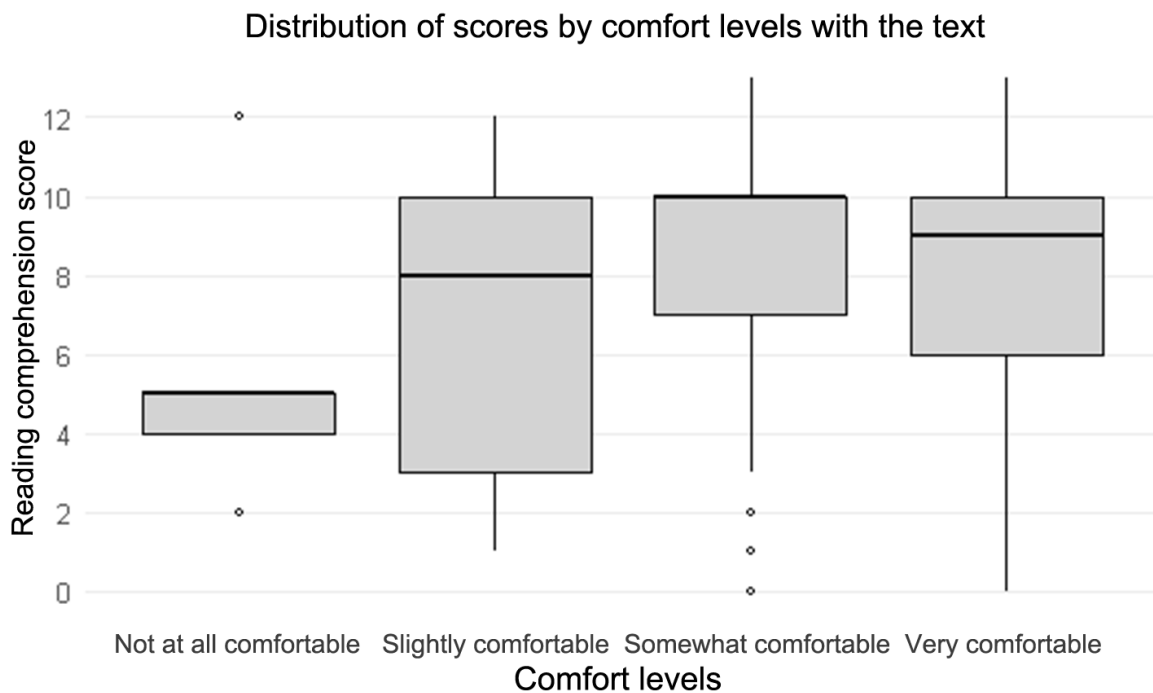


Figure 2. Score distribution by text comfort perception.

Note: The figure displays the distribution of reading comprehension scores according to participants' self-reported comfort levels with the text.

Contrary to expectations from affective-motivational reading theories (McNamara & Magliano, 2009; Schiefele, 2009), students who reported feeling “not at all comfortable” showed higher median comprehension scores and lower score dispersion compared to those who felt more comfortable. The results indicate that the feeling of discomfort is not always linearly associated with poor comprehension outcomes. In some cases, students who reported difficulties actually appeared to engage in more deliberate and attentive reading strategies, attempting to compensate for the challenge. This interpretation is consistent with cognitive perspectives that highlight how moderate levels of textual difficulty can stimulate effortful processing and support comprehension (D’Mello & Graesser, 2012; Kintsch, 1998).

On the other hand, those who described themselves as “very comfortable” or “somewhat comfortable” showed wider variation in their performance. Such dispersion could be explained by differences in strategy use, fluctuating levels of attention, or a degree of overconfidence that reduced the cognitive effort devoted to the task. These observations suggest a tension in the design of academic reading materials: while accessibility is crucial, a certain level of challenge may also be necessary to promote deeper engagement.

Further research is needed to examine how text characteristics, reader profiles, and perceived challenge interact to shape comprehension outcomes.

4.2. Poisson Regression Model

To identify the variables influencing reading comprehension performance, a Poisson regression model was estimated. Predictor variables included academic semester, prior experience with different text types, preferred modality, and perceived comfort level. The final model, selected through stepwise regression and validated with deviance and AIC indices, demonstrated adequate explanatory power (Deviance = 318.7; AIC = 1094.2; $p < 0.001$).

Table 1 presents the estimated coefficients (β), standard errors, z-values, and significance levels. The type of text received was a significant predictor: combined texts led to significantly higher scores compared to verbal-only texts ($\beta = 0.15, p < 0.001$), while graphic-only texts showed a positive but non-significant trend ($\beta = 0.080, p = 0.097$).

Interpretation of coefficients. The results indicate that the type of text was one of the strongest predictors. Students who received combined texts scored significantly higher than those who received only verbal texts ($\beta = 0.15, p < 0.001$), which translates to an expected increase of about 16% in comprehension scores. This advantage highlights the role of dual coding: when verbal and visual resources are integrated coherently, learners can process information through complementary channels, reducing cognitive load and enhancing retention. By contrast, the coefficient for graphic-only texts was smaller and non-significant ($\beta = 0.08, p = 0.097$), suggesting that diagrams without explanatory text may not provide sufficient support, particularly for readers with limited visual literacy.

Table 1. Poisson regression coefficients.

Term	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.36	0.17	13.55	< 0.001***
Text type (Both)	0.15	0.04	3.59	< 0.001***
Text type (Graphic only)	0.08	0.05	1.66	0.097
Semester	-0.04	0.01	-2.75	0.006**
Verbal experience	0.06	0.03	1.99	0.047*
Mathematical experience	-0.09	0.03	-2.84	0.005**
Comfort: Somewhat comfortable	0.01	0.04	0.18	0.861
Comfort: Slightly uncomfortable	-0.19	0.07	-2.49	0.013*
Comfort: Not at all comfortable	-0.39	0.15	-2.65	0.008**
Preference: Graphic/Mathematical only	-0.11	0.07	-1.67	0.096
Preference: Both	0.01	0.06	0.13	0.896
Preference: No preference	0.06	0.08	0.74	0.46

Note: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***) (two-tailed significance levels).

Other significant predictors included.

Academic trajectory also showed an effect. The negative coefficient for semester ($\beta = -0.04$, $p = 0.006$) implies that as students advanced in their academic programs, their expected comprehension scores decreased slightly (around 4% per semester). One interpretation is that more advanced students are exposed to increasingly complex and specialized texts, which may reduce their performance on general comprehension tasks.

Prior experiences further shaped outcomes. Verbal experience exerted a positive effect ($\beta = 0.06$, $p = 0.047$), indicating that students with stronger exposure to verbal texts obtained approximately 6% higher scores. This underscores the continuing importance of traditional literacy skills as a foundation for multimodal comprehension. In contrast, mathematical or graphical experience showed a negative effect ($\beta = -0.09$, $p = 0.005$), associated with a 9% decrease in performance. This finding suggests that familiarity with symbolic or numerical formats does not necessarily translate into better multimodal comprehension and may even hinder performance when texts require integration across multiple modes.

Finally, perceived comfort emerged as a meaningful predictor. Students who reported feeling slightly uncomfortable ($\beta = -0.19$, $p = 0.013$) or not at all comfortable ($\beta = -0.39$, $p = 0.008$) scored 19% and 39% lower, respectively, compared to their more comfortable peers. These results confirm that low comfort is closely tied to poorer comprehension, although the descriptive analysis suggested that some discomfort may trigger compensatory strategies. Taken together, the findings reinforce the need to design reading materials that balance accessibility with a degree of cognitive challenge to foster engagement.

Figure 3 shows the coefficient plot with 95% confidence intervals, providing a visual summary of the direction and strength of each predictor's effect.

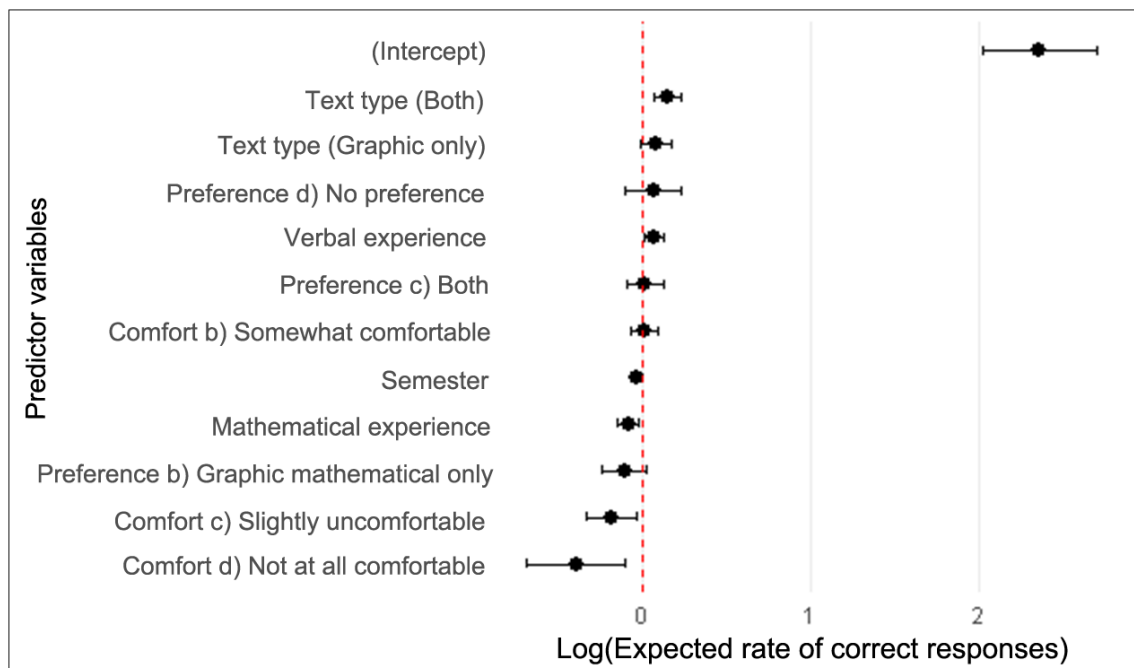


Figure 3. Poisson regression coefficient plot.

Note: Poisson regression estimates with 95% confidence intervals for predictors of multimodal reading comprehension.

While median scores were comparable across modalities, differences in dispersion and outliers were minor but consistent, providing valuable insights into the relative effectiveness of each modality.

Verbal text: Participants in the verbal-only group had a median score of 8, but they produced an even wider range of scores compared to the other groups, 0 to 12. Unlike the "Both" and "Graphic Only" conditions, the verbal-only group did not show extreme outliers, but the wide distribution of scores indicates inconsistent comprehension

outcomes. It seems that some students could work effectively in a linear verbal format, while others were impaired. Ultimately, this trend suggests that verbal-only texts could restrict the potential literacy development of students reading in diverse ways (Kintsch, 1998).

Graphic text: Participants used graphic texts (e.g., infographics and diagrams with limited verbal assistance), and had a median score of 9, with a moderate range and limited low-end outliers. Regardless of the expected strong dispersion, most of the scores were bunched up toward the high end of the distribution; nonetheless, the presence of low values indicates that not all students were equally able to work with and process non-linear visual structures. Overall, this suggests that graphic formats can support comprehension with most students, yet some students (with limited visual literacy or less experience interpreting diagrams) can be disrupted (Bateman, 2014; Mayer, 2014).

Combined text (Verbal + Graphic): Participants in the combined-text conditions had the highest median scores (10), with a relatively simple distribution of scores compared to the other codings, and limited low outliers. Overall, these scores related to the performance group, suggesting the verbal/visual version enhanced for more stable comprehension outcomes, since students could use promptings from both sources. When words and images are coherently integrated, both codings are reiterated, and the cognitive load reduces (Kress & Van Leeuwen, 2006; Mayer, 2014). Although the median did not differ significantly from the other groups, the combined-text group appeared to have an overall more evenly distributed performance across the sample, indicating that multimodal literacies must be regarded as a critical academic capacity.

5. DISCUSSION

A total of 426 undergraduate students; analyses used a Poisson regression framework (Poisson model fit: Deviance = 318.7; AIC = 1094.2; $p < 0.001$). In line with the modality hypothesis, combined texts showed a positive effect ($\beta \approx 0.15$; see Table 1), while one modality condition displayed a positive but non-significant trend ($\beta = 0.080$, $p = 0.097$; Table 1). Consistent with the literacy account, prior verbal experience was associated with higher performance ($\beta = 0.06$, $p = 0.047$; Table 1). Additionally, the semester exhibited a small negative effect ($p=0.006$; Table 1). Figure 2 depicts a graded but non-monotonic relation between comfort and comprehension. The present study examined multimodal reading comprehension among university students from Social Sciences and Engineering, modeling performance as a function of text modality and reader-level variables. Three results are central. First, exposure to coherently integrated verbal-graphic texts predicted higher comprehension scores compared to verbal-only materials. Second, perceived comfort with the text exhibited a graded relationship with outcomes; lower comfort tended to coincide with lower performance, although descriptive distributions also suggested that some discomfort may trigger more deliberate processing in subsets of readers. Third, prior verbal experience was positively associated with performance, whereas self-reported mathematical/graphic experience showed a small but significant negative association.

The modality effect aligns with the core prediction of multimedia learning: when words and images are well aligned in time and space and unnecessary extraneous elements are minimized, learners can distribute processing over partially independent channels, consolidate complementary representations, and reduce the need to infer bridging relations that impose additional load (Mayer, 2009, 2014). In our context, the combined condition was intentionally designed according to contiguity and coherence principles; thus, its advantage is consistent with a dual-channel account without requiring a novelty explanation. In addition, the combined group displayed lower dispersion, a pattern compatible with the hypothesis that multimodal redundancy and cross-cueing stabilize performance across heterogeneous reader profiles by providing multiple entry points to the same underlying content.

The negative association between perceived discomfort and comprehension is unsurprising from a metacognitive perspective low comfort often accompanies high load or weak prior knowledge. Still, the descriptive pattern in which the “not at all comfortable” group attained a relatively concentrated distribution of scores invites a more nuanced reading. One plausible interpretation is that for a subset of students, awareness of difficulty leads to compensatory

strategies (slower reading rate, increased regressions to figures/tables, or explicit checking of intermodal consistency), yielding acceptable performance despite low confidence. Prior research notes that moderate desirable difficulties can elicit deeper processing when design and task demands remain within the learner's zone of proximal development (D'Mello & Graesser, 2012; Kintsch, 1998). In practical terms, discomfort should not be equated with failure; instead, instructors can scaffold strategic behaviors that convert perceived challenges into productive efforts.

The positive effect of prior verbal experience likely reflects the anchoring role of conventional literacy skills. Even in multimodal contexts, readers must still coordinate propositions, evaluate referential cohesion, and integrate local inferences into a coherent situation model (McNamara & Magliano, 2009). More surprising, perhaps, is the small negative coefficient for mathematical/graphic experience. We do not interpret this as evidence against visual literacy; rather, we conjecture that self-reported "graphic experience" is heterogeneous, often emphasizing symbolic manipulation (e.g., algebraic transformations) or diagrammatic conventions that do not automatically transfer to cross-modal alignment. If a reader over-relies on graphs as self-contained answers without synchronizing them with accompanying text or expects a one-to-one mapping that is not present, misalignment costs can accumulate. Prior work on logocentrism (Parodi & Julio, 2017) documents a tendency to privilege a single dominant mode; our data suggest the complementary risk of "graph-centrism" when visuals are processed without textual coordination.

Finally, the small negative semester effect points to a curricular paradox: as students' progress, they likely encounter more specialized and denser materials, but explicit multimodal instruction does not necessarily keep pace. Without targeted training, later semesters can become a moving target in which increased complexity outstrips strategy development. This does not mean higher-semester students "know less"; rather, task difficulty rises faster than their multisemiotic repertoire expands.

5.1. Pedagogical and Curricular Implications

In practice, we recommend short, graded 'cross-reference' prompts. For example: "Using Table 1, identify which predictor best explains the score gap and then verify the pattern in Figure 2." Students must articulate how a textual qualifier (e.g., a rate change) maps onto a visual feature (e.g., a slope or inflection). This explicitly trains the alternation between verbal and graphic anchors that our model estimates capture.

Three levels of implications follow: task design, course pedagogy, and curricular architecture.

(1) Task design. For course materials and assessments, coherence and contiguity are necessary but not sufficient. Instructors should (a) signal intermodal links explicitly (e.g., "In Figure 2, note how the slope corresponds to the rate described in paragraph 3), (b) avoid orphan figures that repeat text verbatim (unnecessary redundancy that consumes resources without adding constraints), and (c) design prompts that require cross-referencing ("According to Table 1 and the preceding paragraph, which interpretation remains consistent?"). Rubrics can include an "intermodal alignment" dimension to make expectations visible. Likewise, when difficulty is intentionally increased to induce effortful processing, designers should bound that difficulty through scaffolds (advance organizers, glossaries of visual conventions, and examples of well-formed text-figure mappings).

(2) Course pedagogy. Two families of interventions are promising. The first is explicit strategy instruction for multimodal integration: teaching students to alternate focal attention between clauses and visual anchors, to verbalize mappings (e.g., "the curve's inflection corresponds to the change described in sentence 4"), and to check whether numerical labels and textual qualifiers mutually constrain interpretation. The second is productive multimodal production: asking students to transform a textual report into an infographic or to accompany a chart with a 120-word micro-explanation that enforces contiguity and economy. Production tasks force readers to reverse-engineer design choices and often reveal latent misunderstandings in ways that pure comprehension tasks do not.

(3) Curricular architecture. A transversal multimodal literacy trajectory is recommended, with discipline-specific branching. For Social Sciences, emphasis should be placed on argument structure, claims-evidence-warrants, and the role of tables in constraining interpretive leaps. For Engineering, emphasis should include conventions for axes, units,

thresholds, and the rhetoric of uncertainty; students should practice translating between diagrams and short expository passages. Across both tracks, mixed-mode labs (reading-to-write and write-to-read cycles) can consolidate patterns: students read a multimodal source, extract structured notes, and then produce a derivative artefact with fidelity and added value.

Institutionally, this implies that multimodal literacy belongs not only to first-year “study skills” courses but also to mid- and late-semester experiences where disciplinary density is maximal. Faculty development should cover basic design principles (layout, hierarchy, alignment, whitespace), common misalignments (figure-text contradictions), and rapid formative diagnostics (two-minute “find and align” checks at the start of class).

5.2. Study Limitations

Our cross-sectional design, single-institution sample, and reliance on self-reports for some predictors delimit generalizability. Because assignment to text type was randomized, internal validity for the modality effect is reasonably strong. However, residual confounds remain possible, e.g., domain familiarity with the statistical topic may vary idiosyncratically across subgroups despite stratification. In addition, the comprehension test, although piloted and reviewed by experts, contains items with varying discrimination; future Rasch modeling could refine the scale and produce person- and item-level diagnostics. Finally, perceived comfort is a broad construct; more granular metacognitive measures (effort ratings, strategy checklists, or eye-tracking indices of regressions between modes) could parse “low comfort but high control” from “low comfort and low control.”

5.3. Directions for Future Research

Four avenues appear most consequential. First, longitudinal designs could track whether targeted instruction narrows dispersion and reduces reliance on a dominant mode (logo-or graph-centrism). Second, experimental manipulations of intermodal signaling (e.g., arrows, callouts, or embedded verbal glosses) could quantify the magnitude of scaffolding benefits and their interaction with prior experience. Third, process-tracing with eye-tracking or click-stream analytics in interactive materials would illuminate time-course dynamics: do high performers exhibit faster and more symmetrical alternation between text and graphics? Fourth, mixed-methods classroom studies could document how instructors actually model intermodal reasoning and whether students internalize a “talk-through” of mappings when producing their own artefacts. Finally, cross-institutional samples in Latin America would separate language- and culture-specific phenomena from general principles of multimodal comprehension.

Extended synthesis. Taken together, our results and the literature converge on a principle of constrained redundancy: reading comprehension benefits when graphics and text neither duplicate nor diverge, but instead restrict each other’s interpretive degrees of freedom. The educational corollary is straightforward: curriculum should prioritize the design and teaching of constraints on how to make a figure “say” something precise and how to make a sentence bind a visual trend to a causal or mechanistic account. In this sense, multimodal literacy is not an accessory technical skill; it is a form of epistemic discipline that lies at the heart of university learning.

6. CONCLUSIONS

This study confirms that multimodal reading comprehension is significantly influenced by text type, reader attitudes, and prior experience with various semiotic modes. In particular, texts that coherently integrate verbal and graphic resources enhance comprehension, consistent with multimedia learning theory.

The prominent role of perceived comfort emphasizes the importance of affective and metacognitive factors in processing multimodal texts. Furthermore, students’ disciplinary profiles influence their ability to engage with different modes, reinforcing the necessity of tailored multimodal training. From an educational standpoint, we

conclude that university curricula should incorporate multimodal literacy as a transversal dimension of academic training, not merely as a technical skill but as a fundamental cognitive, communicative, and critical competence.

Funding: This research was supported by Universidad Católica de Santa María (Grant number: P-2022-024).

Institutional Review Board Statement: The study involved minimal risk and followed ethical guidelines for social science fieldwork. Formal approval from an Institutional Review Board was not required under the policies of Institute for Research Ethics Committee of the Universidad Católica de Santa María, Perú. Informed verbal consent was obtained from all participants, and all data were anonymized to protect participant confidentiality.

Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

Disclosure of AI Use: The author used OpenAI's ChatGPT (GPT-4) to edit and refine the wording of the Introduction and Literature Review. All outputs were thoroughly reviewed and verified by the author.

REFERENCES

- Barrett, T. C. (1976). Taxonomy of reading comprehension. In R. C. Smith (Ed.), *Teaching reading in the middle grades*. In (pp. 51–60). Reading, MA: Addison-Wesley.
- Bateman, J. (2019). *Theories of multimodality: Contemporary perspectives*. New York: Routledge.
- Bateman, J. A. (2014). *Text and image: A critical introduction to the visual/Verbal divide*. New York: Routledge.
- Bezemer, J., & Kress, G. (2016). *Multimodality, learning and communication: A social semiotic frame*. New York: Routledge.
- Cope, B., & Kalantzis, M. (2009). "Multiliteracies": New literacies, new learning. *Pedagogies: An International Journal*, 4(3), 164–195. <https://doi.org/10.1080/15544800903076044>
- D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145–157. <https://doi.org/10.1016/j.learninstruc.2011.10.001>
- Danielsson, K., & Selander, S. (2021). Semiotic modes and representations of knowledge. In *Multimodal texts in disciplinary education: A comprehensive framework*. In (pp. 17–23). Heidelberg, Germany: Springer.
- De-la-Peña, C., Chaves-Yuste, B., & Luque-Rojas, M.-J. (2024). Digital reading comprehension: Multimodal and monomodal inputs under debate. *Reading Psychology*, 45(5), 500–518. <https://doi.org/10.1080/02702711.2024.2339809>
- Gladic-Miralles, J., & Cautín-Epifani, V. (2018). Levels of comprehension and their relationship with the predominance of semiotic systems: An approach to multimodal comprehension from academic discourse. *Estudios Pedagógicos (Valdivia)*, 44(1), 293–313.
- Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education*, 32(1), 241–267. <https://doi.org/10.3102/0091732X07310586>
- Jewitt, C. (2009). *The Routledge handbook of multimodal analysis*. New York: Routledge.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, U.K: Cambridge University Press.
- Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London, UK: Oxford University Press.
- Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). New York: Routledge.
- Martínez Lirola, M., & Ibañez Castejon, L. (2015). *Learning with multimodal texts: A practical experience in master's classes*. In M. T. Tortosa Ybáñez, J. D. Álvarez Teruel, & N. Pellín Buades (Eds.), Paper presented at the 13th Conference on Research Networks in University Teaching: New Organizational and Methodological Strategies in University Education to Respond to the Need for Adaptation and Change.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. In (2nd ed., pp. 43–71). Cambridge, England: Cambridge University Press.

- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, 51, 297-384. [https://doi.org/10.1016/S0079-7421\(09\)51009-2](https://doi.org/10.1016/S0079-7421(09)51009-2)
- Parodi, G. (2015). Variation across university genres in seven disciplines: A corpus-based study on academic written Spanish. *International Journal of Corpus Linguistics*, 20(4), 469-499. <https://doi.org/10.1075/ijcl.20.4.03par>
- Parodi, G., & Julio, C. (2015). Beyond words: Can the Monetary Policy Report discursive genre be understood from a single predominant semiotic system? *ALPHA Journal*, 41, 133-158.
- Parodi, G., & Julio, C. (2017). There are not only words in written texts: Some theories and models of multimodal or multisemiotic text comprehension. *Reading Research*, 8, 27-48.
- Pflaeging, J., Bateman, J. A., & Wildfeuer, J. (2021). Empirical multimodality research: The state of play. In J. Pflaeging, J. Wildfeuer, & J. A. Bateman (Eds.), *Empirical multimodality research: Methods, evaluations, implications*. In (pp. 1-32). Boston, MA: De Gruyter.
- Pinheiro, M. S. (2016). Multimodality and visual literacy in the Spanish language classroom: Analysis of a written production activity. *Brazilian Journal of Applied Linguistics*, 16(4), 575-593. <https://doi.org/10.1590/1984-639820159956>
- Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school*. In (pp. 197-222). New York: Routledge.
- Serafini, F. (2012). Reading multimodal texts in the 21st century. *Research in the Schools*, 19(1), 26-32.
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40-59. <https://doi.org/10.17763/haer.78.1.v62444321p602101>
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Corporation.

Views and opinions expressed in this article are the views and opinions of the author(s), International Journal of Education and Practice shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.