pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

THE EFFECTIVENESS OF MULTIMODAL TEACHING TECHNOLOGIES IN DEVELOPING STUDENTS' SENSORY-PERCEPTUAL COMPETENCIES

Feruz Pardaboyevich Rakhmonkulov
Jizzakh State Pedagogical University associate professor
E-mail: feruzraxmonkulov@gmail.com

Buronov Xolbek Shuxrat ugli

Jizzakh State Pedagogical University, student.
https://doi.org/10.5281/zenodo.20327905

Abstract: This conceptual review examines the effectiveness of multimodal teaching technologies in fostering the sensory-perceptual competencies of university students. Drawing on established theoretical frameworks — the Cognitive Theory of Multimedia Learning, Dual Coding Theory, Cognitive Load Theory, the multimodality tradition in social semiotics, and sociocultural perspectives on mediated learning — the paper synthesises decades of conceptual and empirical work to clarify how multimodal instruction interacts with the perceptual systems through which higher-education learners receive, integrate, and interpret information. A structured narrative-review procedure was applied to seminal monographs and peer-reviewed articles published between 1966 and 2014. The synthesis indicates that well-designed multimodal environments support sensory-perceptual development by (i) distributing processing load across visual and auditory channels, (ii) enabling referential connections between verbal and non-verbal representations, (iii) cultivating perceptual differentiation through guided attention, and (iv) embedding perception in semiotically rich, socially mediated activity. The discussion outlines design implications and identifies boundary conditions — notably the redundancy, split-attention, and modality effects — that constrain when multimodality is beneficial. The paper concludes that multimodal technologies are most effective when their design is principled rather than additive: multiplying channels does not, by itself, multiply learning.

Keywords: multimodal learning, sensory-perceptual competence, dual coding, cognitive load, higher education, instructional design.

Introduction

Higher-education pedagogy has long recognised that learning begins in perception. Before concepts are abstracted, propositions are formed, or arguments are evaluated, learners must first apprehend information through the senses and organise it into structured perceptual wholes. The development of these sensory-perceptual competencies — the capacities to attend selectively, to differentiate features, to integrate information across modalities, and to recognise meaning in patterned input — is therefore a precondition for higher-order learning, not a peripheral concern. As Gibson observed, the perceptual systems are not passive receivers but active, exploratory instruments through which the organism extracts information from a structured environment [11, p. 47].

Contemporary instruction at the university level increasingly relies on technologies that present information in more than one modality simultaneously: animated diagrams accompanied by spoken explanations, interactive simulations combining text and image, video lectures with on-screen captions, and collaborative platforms that integrate gesture, image, and discourse. Such environments are commonly described as multimodal. Kress argues that multimodality is not a stylistic option but a fundamental property of contemporary communication, in which meaning is distributed across image, writing, gesture, and sound rather than concentrated in language alone [8, p. 1]. Educationally, this raises a concrete question: do multimodal teaching

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

technologies, in fact, develop students' sensory-perceptual competencies more effectively than single-modality instruction, and under what conditions?

Three theoretical traditions converge on this question. First, Paivio's Dual Coding Theory holds that cognition employs two functionally independent but interconnected systems — a verbal system specialised for linguistic units and a non-verbal (imagery) system specialised for perceptual analogues — and that referential connections between them enrich representation and retrieval [2, p. 53]. Second, Mayer's Cognitive Theory of Multimedia Learning extends this dual-channel assumption to instructional design, postulating that learners process pictorial and verbal information in separate channels of limited capacity, and that meaningful learning arises when they actively select, organise, and integrate material from both [1, p. 63]. Third, Sweller's Cognitive Load Theory specifies the resource constraints under which such integration can succeed, distinguishing intrinsic, extraneous, and germane sources of load on working memory [4, p. 261].

Alongside these cognitive accounts, sociocultural and social-semiotic perspectives emphasise that perception is itself culturally and historically mediated. Vygotsky argued that higher mental functions, including the organised forms of attention and perception that mature learners deploy, develop through the internalisation of socially mediated activity with cultural tools and signs [7, p. 57]. Jewitt's review of multimodality in school classrooms similarly shows that what counts as competent perception — what is attended to, what is recognised as meaningful — is shaped by the modal resources made available in the learning environment [9, p. 246].

Despite this rich conceptual base, university teachers and instructional designers often face the multimodal turn as a practical, under-theorised problem: more channels are assumed to produce more learning, and richness of stimulation is conflated with quality of instruction. The aim of the present paper is therefore to synthesise the established literature on multimodal instruction with explicit attention to its consequences for sensory-perceptual competence in higher-education learners. The paper asks: (a) Which theoretical mechanisms link multimodal instruction to sensory-perceptual development? (b) What conditions, identified by empirical and conceptual research, make multimodal designs effective or counterproductive? (c) What design implications follow for higher education?

Methods

Because the present article is a conceptual review rather than an empirical study, the methodological work consisted in identifying, appraising, and synthesising the theoretical and empirical literature most directly relevant to the research questions. A structured narrative-review procedure was adopted, comprising four stages.

First, a corpus of foundational sources was established by working backwards from contemporary syntheses to seminal works. The starting points were Mayer's monograph Multimedia Learning [1, p. 63], which anchors the cognitive tradition; Paivio's Mental Representations [2, p. 53] for dual coding; Sweller and colleagues' formulations of cognitive load [4, p. 261; 5, p. 259]; Vygotsky's Mind in Society [7, p. 39] for sociocultural mediation; and Kress's Multimodality [8, p. 79] together with Jewitt's review article [9, p. 246] for the social-semiotic tradition. Gibson's The Senses Considered as Perceptual Systems [11, p. 47] was retained as the principal classical reference on perception itself.

Second, the corpus was extended to include peer-reviewed empirical and theoretical articles that operationalise these frameworks in instructional contexts: Mayer and Moreno on cognitive load in multimedia learning [6, p. 47] and on split-attention and dual processing [14, p. 317]; Moreno and Mayer on interactive multimodal environments [13, p. 318]; Ginns's meta-analysis

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

of the modality effect [12, p. 322]; Baddeley's account of working memory [10, p. 557]; and Paas, Renkl, and Sweller on recent developments in cognitive load theory [3, p. 2].

Third, each source was read against the three research questions and coded for its contribution to (a) the cognitive architecture underlying multimodal processing, (b) empirical evidence on conditions of effectiveness, and (c) implications for the design of higher-education instruction targeting sensory-perceptual competence. Fourth, the coded extracts were synthesised thematically, with attention to convergences across traditions as well as to points of tension — particularly between cognitivist accounts that treat modality as a property of stimuli and semiotic accounts that treat it as a property of socially distributed meaning-making.

No empirical data were generated, and no claims are made that are not traceable to the cited literature. Quantitative effect sizes are reported only where they appear in the original sources.

Results

Cognitive architecture: dual channels and limited capacity. The first and most stable finding of the synthesis is that the human cognitive system is organised in a way that makes multimodal instruction architecturally plausible. Baddeley's model of working memory identifies a phonological loop dedicated to verbal-acoustic information and a visuospatial sketchpad dedicated to visual and spatial information, coordinated by a central executive [10, p. 557]. Paivio's Dual Coding Theory describes an analogous structural separation at the level of long-term representation between verbal and non-verbal (imagery) systems, with referential connections binding the two [2, p. 56]. Mayer formalises this architecture for instructional purposes as the dual-channel assumption: learners process pictorial and verbal material through partially independent channels of limited capacity [1, p. 63]. The implication is that distributing information across visual and auditory channels can, in principle, enlarge the cognitive resources available to a learner — provided the distribution is principled rather than redundant.

The modality effect and its boundary conditions. The empirical claim that follows from the dual-channel assumption is the modality effect: when visual information (such as an animation or diagram) is accompanied by spoken rather than written text, learning is improved because the verbal information is routed through the auditory channel and does not compete with the visual material for limited visual processing. Ginns's meta-analysis of forty-three independent contrasts found a robust overall modality effect favouring audio-visual over visual-only presentations, with average effect sizes that are pedagogically meaningful [12, p. 322]. Mayer and Moreno's split-attention experiments similarly showed that learners who received concurrent narrated animations outperformed those who received the same content with on-screen text, consistent with the dual-processing prediction [14, p. 317].

Crucially, however, the same body of work specifies the conditions under which multimodality fails. The redundancy effect — described by Mayer and Moreno among nine principles for reducing cognitive load — shows that presenting identical information simultaneously as on-screen text and narration can depress learning, because learners attempt to reconcile two streams of identical verbal input [6, p. 47]. The split-attention effect shows that physically separating mutually referring sources (for example, a diagram and its legend on different parts of the screen) imposes extraneous load that erodes learning gains [5, p. 264]. These findings are decisive: multimodality is beneficial when it distributes load across channels and harmful when it duplicates or fragments it.

Referential integration and the depth of perceptual processing. Beyond capacity, multimodal instruction operates on the quality of representation. Paivio argues that when verbal and imagery codes are activated together and referentially linked, the resulting representation is more elaborated, more flexibly retrievable, and more robust against forgetting than a single-code

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

representation [2, p. 67]. Mayer's select-organise-integrate model adds that meaningful learning requires the learner not merely to register both streams but to construct internal verbal and pictorial models and to map them onto each other and onto prior knowledge [1, p. 71]. In sensory-perceptual terms, this is the work of perceptual integration: differentiating features within each modality and binding them across modalities into a coherent percept of the to-be-learned domain. The view is consistent with Gibson's earlier insistence that perception is an active extraction of invariants from structured stimulation, in which the perceptual systems become attuned, through experience, to ever-finer distinctions [11, p. 51].

Cognitive load and the management of intrinsic difficulty. Cognitive Load Theory provides the resource accounting that ties the foregoing claims to design decisions. Sweller distinguishes intrinsic load (the irreducible complexity of the material, given the learner's prior knowledge), extraneous load (the load imposed by the manner of presentation), and germane load (the resources devoted to schema acquisition) [5, p. 259]. Paas, Renkl, and Sweller emphasise that effective instruction reduces extraneous load while permitting germane load to do its work [3, p. 2]. For sensory-perceptual development specifically, this means that the role of multimodal design is not to make tasks easier in a global sense but to reallocate the load so that learners can sustain the perceptual differentiation and integration that the task requires.

Interactivity, agency, and multimodality as a semiotic resource. A second strand of the literature extends the analysis from pre-prepared multimedia to environments in which learners act upon material. Moreno and Mayer's review of interactive multimodal environments concludes that interactivity is most effective when the learner's actions are tied to the instructional content — for example, when learners can manipulate variables and observe consequences — rather than when interactivity is decorative or distracting [13, p. 318]. This converges with the semiotic argument advanced by Kress, who treats each mode (image, writing, gesture, sound) as offering distinct affordances for meaning, such that selecting and combining modes is itself a competence to be developed [8, p. 79]. Jewitt's classroom-based review documents how the modal choices made by teachers shape what is perceptually salient for learners and, in turn, what they come to recognise as meaningful in a discipline [9, p. 252].

Mediation: perception as a culturally formed function. Finally, sociocultural theory situates these mechanisms in a developmental and cultural context. Vygotsky's analysis of the development of higher mental functions describes how attention and perception, initially organised by external stimuli, become voluntarily directed through the mediation of signs and tools provided by the social environment [7, p. 39]. Applied to higher education, the implication is that multimodal technologies are not merely conduits of information but cultural instruments that reorganise what students attend to, how finely they discriminate, and which patterns they treat as meaningful. The development of sensory-perceptual competence in a university discipline — reading an X-ray, hearing a chord progression, parsing a syntactic tree — is precisely the formation of such culturally mediated perception.

Convergence across traditions. Although the cognitive, semiotic, and sociocultural traditions reviewed here originate in different theoretical commitments and rely on different methods, they converge on a small set of claims that bear directly on the present question. All three agree that perception is structured rather than passive: it depends on the organisation of the stimulus, on the active work of the perceiver, and on the historically formed categories that the perceiver brings to the encounter [11, p. 51; 1, p. 71; 7, p. 39]. All three agree that the modality through which information is delivered is consequential and that mismatches between the modality of the stimulus and the cognitive or semiotic resources of the learner degrade learning [5, p. 264; 8, p. 79]. And all three agree, in their own vocabularies, that competence is built through repeated, guided encounters with structured material rather than through one-shot

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

exposure to rich stimulation [2, p. 67; 13, p. 318]. These convergences supply the warrant for treating multimodal teaching technologies as a coherent object of pedagogical analysis, despite the heterogeneity of their theoretical foundations.

Discussion

Taken together, the findings of the synthesis support a qualified affirmative answer to the central question. Multimodal teaching technologies can be effective vehicles for the development of students' sensory-perceptual competencies, but the effectiveness is conditional on design choices that align the instructional surface with the cognitive architecture of the learner and with the semiotic structure of the discipline.

The first design implication concerns channel allocation. Where instruction unavoidably presents visually complex material — anatomical diagrams, statistical graphics, engineering schematics — accompanying it with spoken rather than written explanation is, on the present evidence, the conservative choice [12, p. 322; 14, p. 317]. The visual channel remains free to support perceptual differentiation within the diagram while verbal scaffolding arrives through the auditory channel.

The second implication concerns the avoidance of redundancy and split attention. Presenting identical information as both on-screen text and narration, or distributing labels and their referents across non-adjacent regions of a display, predictably erodes the benefits that the dual-channel architecture would otherwise confer [6, p. 47; 5, p. 264]. Designers should therefore audit multimodal materials not for how many modes are present but for whether each mode contributes non-redundant information that the learner needs in order to construct an integrated model.

The third implication concerns interactivity. The literature consistently distinguishes intrinsic interactivity — where the learner's manipulations alter the represented system in pedagogically informative ways — from extrinsic or decorative interactivity, which tends to dissipate attention without enriching representation [13, p. 318]. In the development of sensory-perceptual competence, the former is particularly valuable because it permits the learner to vary features systematically and to observe the resulting perceptual transformations, supporting Gibsonian perceptual learning.

The fourth implication concerns the disciplinary specificity of perception. The sociocultural and semiotic literatures remind us that the competent perception of a domain is itself a cultural achievement; what is perceptually salient to a clinician, a musician, or a linguist is the product of mediated activity in a community of practice [7, p. 57; 9, p. 252]. Multimodal technologies in higher education should therefore be evaluated not only against domain-general cognitive principles but also against the modal repertoires of the discipline being taught: whether learners are being inducted into seeing, hearing, and reading as members of that discipline do.

The synthesis also surfaces tensions that should be acknowledged. The cognitivist tradition models modality as a property of stimuli and processes; the semiotic tradition models it as a property of socially organised meaning-making. Neither account is sufficient on its own. A cognitively well-designed environment that ignores disciplinary semiotics may produce learners who can decode displays but cannot perceive disciplinary meaning; a semiotically rich environment that ignores cognitive load may overwhelm learners and prevent the very integration it aims to foster. The most defensible position is therefore integrative: multimodal design must satisfy cognitive-architectural constraints while serving disciplinary semiotic ends.

Finally, a methodological caveat applies. The present article is a conceptual review of established literature; it does not produce new empirical estimates of effectiveness. The effect

pISSN:2349-5707 Volume:13,Issue 5, May -2026 eISSN:2349-5715

sizes reported here are those of the original studies and meta-analyses, drawn from contexts that do not always match the specific conditions of a given university course. Practitioners adapting these findings should treat them as principled defaults rather than fixed prescriptions, and evaluate outcomes locally.

Conclusion

Multimodal teaching technologies are effective for the development of sensory-perceptual competencies in university students when their design is grounded in what is now a robust theoretical and empirical literature. The dual-channel architecture of working memory and the referential structure of dual coding establish the cognitive basis for combining visual and verbal information; the modality, redundancy, and split-attention effects specify the conditions under which this combination helps or hurts; and the sociocultural and semiotic traditions ground the entire enterprise in the cultivation of disciplinary perception. The central lesson is one of restraint: adding modes is not the same as adding learning. Multimodality is a resource to be designed, not a virtue to be displayed. Where this restraint is observed, multimodal technologies can do more than transmit information — they can help shape, in higher-education learners, the trained perception on which expert understanding ultimately depends.

References

1. Mayer, R. E. Multimedia Learning. 2nd ed. New York: Cambridge University Press, 2009. 304 p.

2. Paivio, A. Mental Representations: A Dual Coding Approach. New York: Oxford University Press, 1986. 322 p.

3. Paas, F., Renkl, A., Sweller, J. Cognitive load theory and instructional design: Recent developments. Educational Psychologist, 2003, 38(1), pp. 1–4.

4. Sweller, J. Cognitive load during problem solving: Effects on learning. Cognitive Science, 1988, 12(2), pp. 257–285.

5. Sweller, J., van Merriënboer, J. J. G., Paas, F. G. W. C. Cognitive architecture and instructional design. Educational Psychology Review, 1998, 10(3), pp. 251–296.

6. Mayer, R. E., Moreno, R. Nine ways to reduce cognitive load in multimedia learning. Educational Psychologist, 2003, 38(1), pp. 43–52.

7. Vygotsky, L. S. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press, 1978. 159 p.

8. Kress, G. Multimodality: A Social Semiotic Approach to Contemporary Communication. London: Routledge, 2010. 212 p.

9. Jewitt, C. Multimodality and Literacy in School Classrooms. Review of Research in Education, 2008, 32(1), pp. 241–267.

10. Baddeley, A. D. Working memory. Science, 1992, 255(5044), pp. 556–559.

11. Gibson, J. J. The Senses Considered as Perceptual Systems. Boston: Houghton Mifflin, 1966. 335 p.

12. Ginns, P. Meta-analysis of the modality effect. Learning and Instruction, 2005, 15(4), pp. 313–331.

13. Moreno, R., Mayer, R. E. Interactive multimodal learning environments. Educational Psychology Review, 2007, 19(3), pp. 309–326.

14. Mayer, R. E., Moreno, R. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. Journal of Educational Psychology, 1998, 90(2), pp. 312–320.