Copyright © 2026, authors

http://dx.doi.org/10.21622/ILCC.2026.06.2.2412

Vol. 5, Iss. 2, December 2026

Beyond Feedback-Mode Competition: A Critical Integrative Review of Source, Delivery Mode, and Feedback Use in Higher-Education L2 Writing

Maryse Riad[1], Lobna ElKhabaty[2] and Ramy Shabara[3]

[1] The Arab Academy for Science, Technology & Maritime Transport/ American University in Cairo
[2]Royal Roads University, Ras al Khaimah Campus, UAE
[3]University of Hertfordshire, hosted by The Global Academic Foundation, Egypt

E-Mails: **m.m.riad@xed.aucegypt.edu, lnaguib10@acacia.edu, r.shabara@herts.ac.uk**

Received 08 September 2026 Accepted 10 October 2026 Published 14 October 2026

ABSTRACT

Feedback in higher-education L2 writing is shaped by how feedback is delivered and who or what provides it. This critical integrative review synthesizes 23 empirical studies published from 2010 through July 2026 to examine students' and teachers' perceptions and experiences of feedback across written, audio, audiovisual/screencast, multimodal, AI-mediated, and human–AI configurations. The review distinguishes feedback source from delivery mode and treats perception, preference, feedback uptake, revision, and writing performance as related but non-equivalent forms of evidence. Across the literature, no feedback mode emerges as universally superior. Written feedback supports precision and retrievability; audio offers elaboration and interpersonal presence; audiovisual feedback combines explanation with visual localization; and AI-mediated feedback offers immediacy and scalability. However, these affordances are conditioned by learner characteristics, task and genre, feedback purpose and timing, accessibility, feedback literacy, and teacher workload. Students and teachers appear to value clarity, specificity, personalization, and explanatory richness, although teachers additionally emphasize feasibility and sustainability. AI evidence is similarly heterogeneous, with current research more supportive of human–AI complementarity than substitution. The review's central contribution is a shift from feedback-mode competition to a configurational account in which source, delivery mode, context, stakeholder experience, and feedback use are understood as interdependent. The proposed Complementary Feedback Ecology Model organizes these relationships heuristically and identifies priorities for direct stakeholder comparisons, workload-sensitive designs, and a more rigorously specified feedback research agenda. The findings provide new insights into how feedback configurations can align pedagogical purpose, learner needs, and sustainable implementation in higher-education writing.

Keywords: *Feedback; L2 Writing; Perceptions; Critical Integrative Review*

1. INTRODUCTION

Writing in a second or additional language requires learners to coordinate linguistic accuracy, lexical choice, organization, argumentation, and rhetorical purpose (Matikainen 2024). Feedback is therefore central to writing instruction because it can direct learners' attention to problems and inform subsequent revision (Sasaki and Takeuchi 2026). Yet feedback is shaped not only by what is communicated but also by how it is delivered and by who or what provides it.

Research has expanded from written comments to audio and audiovisual/screencast feedback, with evidence pointing to distinctive rather than universally superior affordances. Audio can provide greater elaboration, personalization, and interpersonal presence, but may be less retrievable and more demanding to produce (Lunt and Curran 2010; Voelkel and Mello 2014; Nemec and Dintzner 2016). Audiovisual feedback can combine explanation with visual localization and, in some settings, support feedback integration, although

positive perceptions do not consistently translate into stronger revision outcomes (Ali 2016; Özkul and Ortaçtepe 2017; Cunningham 2019; Cavaleri et al. 2019).

Comparative evidence, therefore, does not support a straightforward progression from written to audio to audiovisual feedback. Mode effects vary with learner characteristics, task and genre, feedback purpose, accessibility, and implementation conditions; multiple modes may also be valued for their complementarity (Bakla 2020; Ryan, Henderson, and Phillips 2019). This pattern suggests that the more productive question is not which mode is best, but how different feedback resources function under particular conditions.

Generative AI introduces an additional conceptual distinction because it is primarily a feedback source rather than a delivery mode. AI can generate or support feedback delivered through textual or multimodal interfaces, and recent studies report heterogeneous outcomes across AI, teacher, and hybrid configurations (Khojasteh, Soori, and Javed 2025; Alnemrat et al. 2025). Recent reviews and meta-analytic evidence likewise indicate modest but heterogeneous effects and support a supplementary, human-in-the-loop role for AI (ElEbyary, Shabara, and Boraie 2024; Karagöz 2025; Liu and Liang 2026). Taken together, the literature points toward feedback as a configuration of source, mode, context, stakeholder experience, and use rather than a competition among media.

1.1 *The Research Gap*

The central gap is not an absence of research on feedback perceptions but a fragmentation and conceptual heterogeneity of evidence across feedback source, delivery mode, stakeholder group, context, and outcome. First, source and mode are often conflated, obscuring how the same source can operate through different modes and how the same mode can carry feedback from different sources. For example, studies comparing "teacher-written" and "teacher audio" feedback examine different delivery modes while holding the feedback source constant (e.g., Alharbi and Alghammas 2021). Conversely, studies of "screencast feedback" may examine a multimodal delivery configuration that combines written annotations, visual highlighting, and oral commentary rather than a single mode of feedback (e.g., Ali 2016; Bush 2020). These examples indicate why feedback source and delivery mode require separate analytical treatment: the same source can operate through different modes, while a single delivery configuration can incorporate multiple modalities. Without this distinction, comparisons across studies risk treating heterogeneous feedback interventions as equivalent categories. Second, student and teacher evidence are unevenly distributed: although some studies examine both groups (e.g., Cavanaugh and Song 2014), most report one stakeholder perspective, making apparent convergence across studies inferential rather than directly demonstrated. Third, perceptions are not consistently connected to actual feedback use, revision, or writing development. Positive evaluations, therefore, cannot be assumed to indicate greater uptake or learning.

An integrative review is well-suited to this problem because it can synthesize heterogeneous empirical evidence while critically integrating complementary and competing perspectives to generate a more coherent account (Cronin and George 2023; Torraco 2005). The present review, therefore, moves beyond asking which mode is preferred and examines how feedback source and delivery mode interact with learner, task, and instructional conditions to shape stakeholder experience and feedback use.

1.2 *Purpose and Research Questions*

The purpose of this critical integrative review is to synthesize empirical evidence on how feedback source and delivery mode relate to stakeholder experience, feedback use, and writing-related outcomes in higher-education L2 writing. It integrates evidence on established and emerging feedback configurations to identify conditional patterns and develop a heuristic account of how source, mode, context, and use are related.

RQ1. What affordances and constraints of written, audio, audiovisual/screencast, and multimodal feedback are reported in higher-education writing research, and how are these perceived by students and teachers?

RQ2. How do student and teacher perceptions converge and diverge across feedback sources and delivery modes, and what conditions appear to shape these patterns?

RQ3. What does the emerging literature on AI-mediated and human–AI hybrid feedback reveal about source, usefulness, pedagogical appropriateness, and the relationship between perceived value and writing outcomes?

RQ4. What integrative framework can account for the relationships among feedback source, delivery mode, contextual conditions, stakeholder experience, and feedback use?

2. METHOD

2.1 *Review Design and Analytical Logic*

This study adopts a critical integrative review design because the evidence base comprises heterogeneous

qualitative, survey, mixed-methods, experimental, quasi-experimental, and large-scale studies that cannot be meaningfully reduced to a single pooled estimate. Following Cronin and George's (2023) emphasis on fair representation and synthesis across research conversations, the review used the following analytical flow: (a) identify and verify relevant empirical studies; (b) extract study-level information on source, mode, context, stakeholders, affordances, constraints, and outcomes; (c) distinguish perception, preference, feedback uptake, and writing-performance evidence; (d) compare patterns within and across studies; and (e) integrate the findings into a conceptual framework. The review is purposive and citation-chained rather than a reproducible systematic review; accordingly, it makes no claim of exhaustive literature identification.

2.2 Scope and Evidence Identification

The review covered empirical research published from 2010 through July 2026 in higher-education writing contexts, with particular emphasis on L2, EFL, ESL, English-medium, and multilingual writing. The evidence-identification process was retrospectively reconstructed from the search and verification procedures used to assemble the final evidence base. Because the review was conceived as a critical integrative review rather than a registered systematic review, the search was purposive and iterative rather than designed to claim exhaustive database coverage.

The identification process combined database searching, targeted publisher and DOI searches, repository verification, and backward and forward citation chaining. ERIC was used as the principal bibliographic database, while publisher platforms, DOI records, and institutional or subject repositories were used to verify bibliographic information and retrieve potentially relevant studies. Citation chaining was subsequently used to identify empirical studies that were not captured through the initial database and targeted searches.

The search concepts were developed around three clusters: (a) higher-education/L2 writing, (b) feedback source and delivery mode, and (c) stakeholder experience and feedback use. A reconstructed core search expression was:

> ("higher education" OR universit* OR college* OR tertiary OR undergraduate*) AND ("L2 writing" OR "second language writing" OR "EFL writing" OR "ESL writing" OR "academic writing") AND (feedback OR "teacher feedback" OR "written feedback" OR "audio feedback" OR "video feedback" OR "screencast feedback" OR "audiovisual feedback" OR "AI feedback" OR "ChatGPT" OR "automated feedback" OR "hybrid feedback") AND (perception* OR preference* OR experience* OR usability OR workload OR uptake OR engagement OR revision OR "writing performance")

Because database interfaces and indexing conventions differ, equivalent combinations of these concepts were used when conducting targeted searches and verification. Searches and subsequent citation-chaining were restricted to publications falling within the review period of 2010 through July 2026.

The identification process was iterative. Candidate studies were first screened against the substantive scope of the review, after which their reference lists and citing publications were examined to identify additional empirical evidence. Records were then verified against publisher information, DOI records, ERIC entries, and institutional or subject repositories where necessary. This procedure was particularly important for the rapidly developing AI-feedback literature, where terminology varies considerably across studies.

Importantly, the present account is a retrospective reconstruction of the identification strategy rather than a claim that an exhaustive, reproducible, systematic search was conducted contemporaneously. The review, therefore, makes no claim that the 23 studies constitute the complete population of relevant publications. Rather, they constitute the empirically relevant evidence base identified, verified, and retained for the purposes of the integrative synthesis.

2.3 Eligibility Principles

Candidate records were screened against predefined eligibility criteria. Studies were included if they (a) examined feedback in higher-education writing contexts; (b) involved university students, instructors, or both; (c) investigated teacher, digital, audiovisual, AI-mediated, or human–AI/hybrid feedback; (d) reported empirical evidence on perceptions, preferences, experiences, usability, workload, engagement, feedback uptake, revision, writing performance, or related outcomes; (e) provided sufficient bibliographic information for verification; and (f) were published between 2010 and July 2026. Studies were excluded if they focused exclusively on primary or secondary education, addressed feedback unrelated to writing, were purely conceptual without relevant empirical evidence, could not be bibliographically verified, or fell outside the review questions and defined feedback scope.

Systematic reviews and meta-analyses were consulted to contextualize the field and cross-check interpretations, particularly in the AI-feedback literature, but were not treated as primary evidence.

Several safeguards strengthened the credibility of the evidence base. Peer-reviewed journal articles and

indexed proceedings were prioritized, and bibliographic and factual information was verified against publisher, DOI, ERIC, and repository records. Key methodological characteristics, including sample size, design, data sources, and response rates where applicable, were documented. Interpretive weight was given primarily to patterns corroborated across independent studies, while findings from single or small-scale studies were considered cautiously alongside their stated limitations. These procedures supported a structured assessment of the evidence but did not constitute a formal risk-of-bias or methodological quality appraisal.

The final evidence base comprised 23 empirical studies published between 2010 and July 2026. Given substantial variation in study designs, participants, feedback configurations, outcomes, and contexts, the evidence was synthesized qualitatively rather than statistically aggregated. Because a contemporaneous record of all database retrievals and exclusion decisions was not retained, the review does not reconstruct numerical screening counts or claim a PRISMA-style exhaustive search. Instead, the eligibility criteria, verification procedures, and substantive selection boundaries are reported transparently to clarify the basis of the 23-study synthesis.

2.4 *Data Analysis and Characteristics of the Evidence Base*

The data collection process resulted in 23 studies covering diverse geographic settings, including the United Kingdom, United States, Turkey, Saudi Arabia, Malaysia, Lebanon, Australia, Jordan, Iran, Vietnam, China, and Egypt. Each study was examined for educational context, participant group, feedback source, and delivery mode, research design, perception dimensions, reported affordances and constraints, evidence of feedback uptake or revision, teacher workload or feasibility, and hybrid feedback. The synthesis distinguished perceptions, preferences, actual feedback use or uptake, and writing-performance outcomes, which were not treated as interchangeable. For student–teacher comparisons, convergence or divergence was classified as either directly demonstrated within a study or inferred across studies.

The evidence base is heterogeneous in design and scale. Accordingly, findings were synthesized thematically rather than reduced to a single ranking of feedback modes or pooled effect. The literature also reflects a shift from early comparisons of audio and written feedback toward audiovisual/screencast feedback, multimodal combinations, and, since 2024, AI-mediated feedback.

Table 1 summarizes representative empirical studies by context and participants, feedback configuration, and principal evidence; the table illustrates both the breadth of the evidence base and the substantial variation in research design and outcome type.

Table 1: Representative Empirical Evidence on Feedback Configurations

Study	Context / Participants	Design	Feedback configuration	Main finding	Key limitation relevant to interpretation
Lunt and Curran (2010)	UK HE; 60 students; 26 survey respondents	Comparative intervention/student survey	Audio vs. written	Audio feedback was generally positively received and was associated with perceptions of greater detail and engagement.	Small respondent sample and emphasis on student perceptions limit conclusions about longer-term writing development.
Brearley and Cullen (2012)	UK HE; 40 students	Formative classroom-based study	Formative audio feedback	Audio feedback was perceived as useful for providing timely and constructive formative guidance.	Single cohort and formative context limit generalizability; evidence is primarily experiential rather than comparative.
Cavanaugh and Song (2014)	US online composition: students and instructors	Comparative mixed stakeholder study	Audio vs. written	Students and instructors reported different advantages and concerns concerning audio feedback, including personalization and workload.	Context-specific online composition setting; stakeholder comparisons do not establish universal preferences.
Voelkel and Mello (2014)	UK HE	Comparative study	Audio vs. written	Audio increased the amount/richness of feedback in some respects, but greater satisfaction did not necessarily translate into stronger attainment.	Limited setting and outcome measures; perception and attainment were not equivalent constructs.

Ali (2016)	Egypt, 63 university EFL students	Mixed-methods, pre/post comparison	Screencast + written vs. written	Screencast feedback was associated with stronger higher-order writing outcomes and positive perceptions; technical access problems were reported.	Small, context-specific sample; the experimental condition combined video and written feedback, complicating attribution to the screencast alone.
Nemec and Dintzner (2016)	Pharmacy HE	Comparative feedback study + survey	Audio vs. written	Audio feedback contained more feedback and was perceived as personalized, but required greater teacher production time.	Discipline-specific setting and retrospective/matched feedback comparison limit generalization.
Özkul and Ortaçtepe (2017)	Turkey: university EFL writing	5-week experimental study	Video vs. written/correction codes	Video feedback resulted in greater incorporation/correction in subsequent drafts and was positively perceived.	Short intervention and institution-specific context; perception data came from the experimental group only.
Keane, McCrea, and Russell (2018)	US undergraduate students	Survey-based study	Voice + written	66.7% preferred the addition of voice comments to written feedback.	Preference-focused survey; preference cannot establish effects on revision or writing development.
Cavaleri et al. (2019)	Australian HE	Feedback/revision analysis + interviews	Audiovisual vs. written	Audiovisual feedback was associated with successful engagement with feedback and higher odds of successful revision.	Very small interview component (n=3) and institution-specific academic-language-support context.
Cunningham (2019)	US intermediate ESL writing; 12 students	Pre-experimental convergent mixed-methods; crossover	Screencast vs. text	Students preferred screencast for clarity, efficiency, and ease of use; successful revision rates were similar across modes.	A very small purposeful sample and an exploratory within-group design constrain causal and generalizability claims.
Ryan, Henderson, and Phillips (2019)	Australia: 4,514 university students	Large-scale online survey	Multiple digital/non-digital modes	Digital recordings and electronic annotations were highly rated; multiple modes were associated with perceptions of greater detail, personalization, and usability.	Cross-sectional self-report data and heterogeneous feedback experiences prevent causal conclusions.
Bakla (2020)	Turkey; 33 intermediate Turkish-L1 EFL learners	Embedded mixed-methods, multi-draft comparison	Written, audio, screencast	Audio produced the highest number of correct revisions in one task, whereas no significant mode difference occurred in another; students reported benefits and drawbacks for all modes.	A small sample and task-specific variation demonstrate why findings cannot be generalized as a stable mode hierarchy.
Abdu and Mohammed (2021)	Saudi Arabia, university EFL	Comparative perception/revision study	Oral, written, audio, screencast	Preferences were shaped by comprehensibility, specificity, multimodality, and contextual considerations.	Relatively context-bound evidence and multiple feedback dimensions make attribution to mode alone difficult.
Alharbi (2022)	Saudi Arabia; 30 EFL learners	Comparative intervention/revision study	Oral, text, audio, audiovisual	Audiovisual feedback showed strong incorporation/integration in students' revisions.	A small sample and short-term revision focus limit conclusions about sustained writing development.
Roy and Vetter (2023)	US multilingual composition; 8 focal students	Qualitative interview + draft-comparison study	Audio vs. written	Preferences were related to listening confidence, self-efficacy, and students' experiences of the two modes.	Eight focal participants and a qualitative design provide depth but limited transferability.

Cao (2023)	Vietnam; 89 EFL undergraduates	Perception-focused empirical study	Screencast	Students generally viewed screencast feedback positively but identified time and information-load concerns.	Perception evidence provides limited information about actual revision or writing gains.
Saeed, Abu-Sa'aleek, and Alharbi (2024)	Saudi EFL writing	Linguistic/discourse analysis	Written, audio, screencast	Feedback mode influenced evaluative language and teacher positioning.	Primarily analyzes feedback discourse rather than learner uptake or writing development.
Zeevy-Solovey (2024)	EFL writing learners	Comparative perception study	Peer, ChatGPT, teacher feedback	Students differentiated among feedback sources and expressed distinct perceptions/preferences.	Perception-focused comparison; does not establish whether perceived preferences resulted in superior revision or learning.
Han and Li (2024)	China: 102 students, 4 instructors	Classroom mixed-methods intervention	ChatGPT-supported teacher feedback	ChatGPT-assisted teacher feedback broadened the range of feedback and supported deeper engagement/uptake.	Hybrid design makes the independent contribution of ChatGPT difficult to isolate; context involves a specific teacher-mediated implementation.
Alnemrat et al. (2025)	Jordan: undergraduate EFL learners	Quasi-experimental pretest–posttest factorial design	AI vs. teacher	Both groups improved; no statistically significant difference between AI and teacher feedback was observed.	Single argumentative-writing context and structured implementation limit transferability to other genres and feedback conditions.
Khojasteh, Soori, and Javed (2025)	Iran; 88 intermediate EFL learners	Randomized controlled trial	Teacher, AI, hybrid	Hybrid feedback produced the strongest writing gains in the study.	Limited proficiency range and institutional context; the particular AI tools and implementation procedures constrain generalization.
Oanh and Ngoc (2026)	Vietnam, students and teachers	Stakeholder perception study	AI-based writing assessment/feedback	AI was associated with promptness and linguistic support, alongside concerns about depth and the continuing need for human intervention.	Primarily, perception-oriented evidence and rapidly changing AI context limit claims about durable learning effects.
Han and Wang (2026)	China; 35 university EFL learners	Mixed-methods comparative study	Teacher vs. ChatGPT	ChatGPT-supported revisions produced higher revised-essay scores, whereas students rated teacher feedback as more useful and easier to use.	Small sample and single institutional context; discrepancy between performance and perception underscores the need not to equate these outcomes.

Note. HE = higher education.

Table 1 reveals three patterns that are consequential for interpreting the subsequent synthesis. First, the evidence base differs substantially in what it treats as an outcome: some studies examine perceptions or preferences, whereas others examine feedback uptake, revision, writing performance, or implementation. These outcomes cannot be treated as interchangeable indicators of effectiveness. For example, Ryan, Henderson, and Phillips (2019) found strong perceptions of detail, personalization, and usability for several digital configurations, whereas Cunningham (2019) found positive perceptions of screencast feedback without higher revision success. Second, apparent differences between modes are often embedded in different tasks and implementation conditions. Bakla (2020), for instance, found an advantage for audio in one revision task but no significant difference among modes in another, while Cavaleri et al. (2019) reported higher odds of successful revision for audiovisual feedback relative to written feedback. The contrast does not necessarily represent contradictory evidence; rather, it indicates that the relationship between mode and outcome is sensitive to what is being measured and to the conditions under which feedback is used. Third, the table shows that evidence for source effects and evidence for delivery-mode effects are increasingly

intertwined in the recent AI literature. Studies comparing teacher, AI, and hybrid configurations, therefore, cannot be interpreted simply as extensions of written-versus-audio comparisons because the source of feedback changes alongside, or independently of, its delivery form. These patterns justify the review's configurational approach and provide the empirical basis for examining source, mode, context, stakeholder experience, and feedback use as related but analytically distinct dimensions.

3. RESULTS

3.1 *RQ1: Affordances and Constraints of Feedback Delivery Modes*

Written feedback remains valuable for precision, permanence, and retrievability. Its textual form allows learners to return to a specific annotation, locate a linguistic problem, and compare a comment with the relevant portion of a text. However, written feedback can be less interpersonal and may be difficult to interpret when comments are vague or insufficiently contextualized (Abdu and Mohammed 2021; Roy and Vetter 2023).

Audio feedback offers elaboration, conversationalism, personalization, and interpersonal presence. Students often respond positively to its richer explanatory and relational qualities (Lunt and Curran 2010; Voelkel and Mello 2014; Keane, McCrea, and Russell 2018). These benefits are qualified by weaker retrievability, listening demands, technical access, and production time. Nemec and Dintzner (2016), for example, found greater feedback content and personalization but greater faculty production time.

Audiovisual/screencast feedback combines verbal explanation with visual localization and can therefore make the rationale for a revision more transparent. Evidence indicates positive perceptions and, in some contexts, stronger feedback integration or revision (Ali 2016; Özkul and Ortaçtepe 2017; Cavaleri et al. 2019; Alharbi 2022). However, Cunningham (2019) found similar revision success across video and text conditions despite positive perceptions of video. Thus, audiovisual feedback has a distinctive affordance profile but no universal performance advantage.

Across modes, effects are conditional. Bakla (2020) provides a particularly clear illustration of why feedback-mode effects should be interpreted in relation to task demands rather than treated as stable properties of a medium. Across 33 intermediate Turkish-L1 EFL learners, audio feedback produced the highest number of correct revisions in the multi-draft essay-writing task, whereas no significant differences among written, audio, and screencast feedback emerged in the separate essay-revision task. The contrast is important because the same learners encountered the three modes under different task conditions, suggesting that the utility of a mode may depend partly on what the learner is required to do with the feedback. Moreover, participants did not converge on a single preferred mode; their qualitative responses identified advantages and drawbacks associated with each format. Thus, Bakla's findings do not demonstrate that audio is inherently superior. Rather, they show that an advantage observed under one task condition may disappear when the revision activity changes.

This pattern is consistent with the broader evidence. Ryan, Henderson, and Phillips (2019), using a much larger sample of 4,514 university students, found that digital recordings and electronic annotations were highly rated for detail, personalization, and usability, and that students receiving multiple modes were more likely to rate feedback positively on these dimensions. However, this was perception evidence rather than evidence of superior revision or learning. Cunningham (2019) similarly found that students preferred screencast feedback for clarity, ease of use, and efficiency, while successful revision rates remained similar across screencast and text conditions. Taken together, these studies suggest that the affordances of a mode and its effects on writing are related but not equivalent: a mode may improve the accessibility, richness, or perceived usefulness of feedback without necessarily producing a corresponding gain in revision performance.

3.2 *RQ2: Student and Teacher Perceptions*

Students do not show a stable preference hierarchy. They tend to value written feedback for precision and retrievability, audio for personalization and elaboration, and audiovisual feedback for combining explanation with visual localization. Multiple modes can be positively evaluated, but positive perception does not necessarily translate into stronger revision (Ryan, Henderson, and Phillips 2019; Cunningham 2019). Student responses are also shaped by accessibility, listening confidence, prior experience, task demands, and feedback purpose.

Teachers similarly recognize the explanatory and interpersonal advantages of audio and audiovisual feedback but evaluate modes through additional considerations of workload, technical reliability, class size, scalability, and pedagogical fit (Cavanaugh and Song 2014; Nemec and Dintzner 2016). Teachers, therefore, introduce a sustainability dimension that is less visible in student-preference studies.

Table 2 summarizes the principal areas of student–teacher convergence and divergence. It should therefore be read as a map of recurring dimensions rather than as evidence of consensus. The apparent convergence around clarity, specificity, personalization,

and explanatory richness reflects similar values reported in separate student- and teacher-oriented studies, but it does not establish that students and teachers would evaluate the same feedback configuration in the same way. The clearest divergence concerns implementation: students primarily encounter the usability and informational consequences of a feedback mode, whereas teachers must also absorb its production costs, technical demands, and scalability. Nemec and Dintzner's (2016) finding that audio contained more feedback while raising production considerations illustrates this distinction. The table consequently supports a more qualified interpretation of stakeholder alignment: students and teachers may value similar feedback qualities while attaching different practical costs and priorities to achieving them.

Table 2: Student–Teacher Convergence and Divergence Across Feedback Dimensions

Dimension	Student Perspective	Teacher Perspective	Interpretation
Precision	Values written/localized feedback	Values clarity and specificity	Apparent convergence
Personalization	Often values audio/video	Recognizes interpersonal potential	Apparent convergence
Rich explanation	Values audio/audiovisual	Values the ability to elaborate	Apparent convergence
Retrievability	Important for written feedback	Important for practical use	Apparent convergence
Production time	Often less visible to students	Major consideration	Directly studied in Nemec and Dintzner (2016)
Technical access	Can constrain experience	Can constrain implementation	Apparent convergence
Multiple modes	Often positively evaluated	Potentially pedagogically attractive	Directly studied in Ryan, Henderson, and Phillips (2019)
AI immediacy	Potentially attractive	Potentially useful for workload	Partial/inferred convergence
AI contextual judgment	Concerns about genericity	Concerns about accuracy/context	Directly studied in Han and Wang (2026)
Human feedback	Valued for clarity/support	Central to pedagogical judgment	Directly studied in Han and Wang (2026); Nazaretsky et al. (2024)
AI replacing the teacher	Little evidence of demand for this	Caution strongly warranted	Apparent convergence toward supplementation

Note. Rows identified as "directly studied" reflect findings from studies examining both stakeholder groups or comparable perspectives on the same dimension. Rows identified as "apparent convergence" reflect interpretive synthesis across separately reported student- and teacher-side findings and should not be interpreted as direct agreement statistics.

3.3 *RQ3: AI-Mediated and Human–AI Hybrid Feedback*

AI introduces a source dimension rather than simply another delivery mode. In this review, AI-mediated feedback refers broadly to feedback processes involving AI in the generation, delivery, or interaction of feedback. AI-generated feedback is produced primarily by an AI system in response to a learner's text. AI-assisted feedback involves AI supporting a human provider, for example, by identifying problems or suggesting revisions that the teacher evaluates and adapts. Human–AI hybrid feedback combines substantive AI input with human pedagogical judgment in the feedback ultimately received by the learner. Thus, these terms distinguish the broader process (AI-mediated), primary source (AI-generated), supporting role (AI-assisted), and combined human–AI configuration (hybrid).

AI-mediated feedback can therefore take different delivery forms, including written/textual feedback, conversational interaction, or multimodal feedback. Current evidence indicates potential benefits in immediacy, scalability, linguistic analysis, and revision support, but results are heterogeneous. Hybrid teacher–AI configurations have produced particularly promising outcomes, whereas other studies report no significant difference between AI and teacher feedback (Alnemrat et al. 2025; Han and Li 2024; Khojasteh, Soori, and Javed 2025).

Importantly, perceived usefulness and performance do not always align. Han and Wang (2026) found higher revised-essay scores following ChatGPT feedback, while students rated teacher feedback as more useful and easier to use. Nazaretsky et al. (2024) likewise showed that perceptions changed when the presumed

source of feedback was disclosed. The AI evidence, therefore, reinforces the need to distinguish source, perception, uptake, and performance rather than treating "AI feedback" as a single intervention.

The broader evidence points toward human–AI complementarity rather than substitution. Liu and Liang's (2026) meta-analysis found a modest overall positive effect of AI-assisted feedback but substantial heterogeneity, with stronger effects for emotional than cognitive outcomes. Karagöz (2025) similarly identified benefits in speed and linguistic support alongside continuing limitations in contextual and nuanced judgment. AI can expand feedback capacity, while teachers remain important for contextual interpretation, prioritization, rhetorical judgment, and interpersonal support.

3.4 *RQ4: From Feedback-Mode Competition to Feedback Ecology*

The cross-study synthesis indicates that the question "Which feedback mode is best?" is poorly specified. Written, audio, audiovisual, and AI-mediated feedback offer different functional affordances, but these affordances are shaped by learner characteristics, task and genre, feedback purpose and timing, digital access, feedback literacy, and teacher workload. Figure 1 organizes this synthesis by distinguishing feedback source and delivery mode and showing how their interaction with contextual conditions can shape stakeholder experience and feedback use.

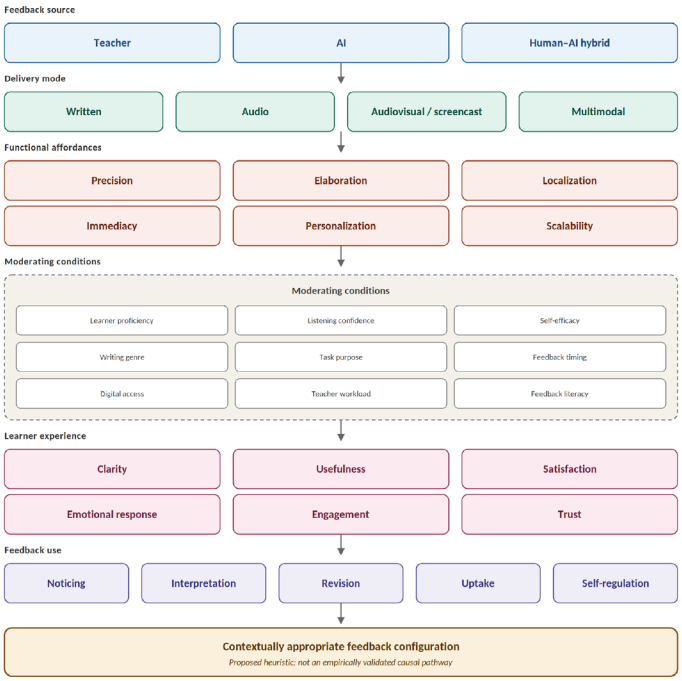


Figure 1: Complementary Feedback Ecology Model

Note. Solid arrows represent relationships synthesized across the reviewed studies; dashed arrows indicate moderating or iterative relationships. The model is heuristic rather than causal. "Ecology" refers to the interacting sources, modes, agents, affordances, constraints, and contextual conditions through which feedback is experienced and used.

Figure 1 displays the proposed relationship among feedback source, delivery mode, functional affordances, contextual conditions, stakeholder experience, feedback use, and outcomes. The model is not a validated causal chain. Rather, it integrates separately reported findings to make visible the central proposition of this review: feedback value emerges from the fit among source, mode, learner, task, purpose, and context, while evidence from use and feasibility can inform subsequent feedback design.

4. DISCUSSION

4.1 *Feedback Effectiveness is Configurational, Not Medium-Inherent*

The most consistent finding is that no context-independent hierarchy of feedback modes is supported. Differences across studies indicate conditional utility rather than simple inconsistency: a mode may be preferred without producing superior revision, and the same mode may function differently across learners, tasks, and purposes. The evidence, therefore, supports a shift from medium competition toward feedback configuration.

4.2 *Perception, Preference, Uptake, and Performance are Distinct*

A central methodological insight is that perception, preference, actual feedback use, revision, and writing performance should not be treated as interchangeable outcomes. Voelkel and Mello (2014) found greater satisfaction with audio without a progression or attainment advantage; Cunningham (2019) found positive video perceptions without higher revision success; and Han and Wang (2026) found higher revised-essay scores following ChatGPT feedback, while students preferred teacher feedback. The review, therefore, supports designs that measure these constructs separately and examine their relationships.

4.3 *Sustainability is Part of Feedback Effectiveness*

Feedback that students value but teachers cannot sustainably produce may not constitute a viable instructional solution. The evidence, however, does not support a simple assumption that richer feedback is necessarily more time-consuming. Nemec and Dintzner (2016) found that audio feedback contained more feedback and was perceived as more personal,

but its production implications differed from written feedback; by contrast, Cunningham (2019) reported that screencast feedback took less time to produce than text feedback in the study's specific implementation. These contrasting findings indicate that workload is not an inherent property of a feedback mode. Rather, it is likely to depend on how feedback is produced, the number and length of comments, technological familiarity, class size, and the particular implementation. Workload should therefore be treated as a contextual condition that can shape the feasibility and sustainability of feedback configurations rather than as a fixed disadvantage of audio or audiovisual feedback.

4.4 *Human–AI Complementarity is More Defensible than Substitution*

AI changes the architecture of feedback production by introducing an additional agent. Current evidence supports its use for speed, coverage, and preliminary linguistic analysis, but also identifies limitations in contextual, rhetorical, genre-specific, and interpersonal judgment. The most defensible interpretation is therefore human–AI complementarity: AI can extend feedback capacity while teachers retain responsibility for contextualization, prioritization, verification, and pedagogical judgment.

4.5 *From Feedback Configuration to Feedback Ecology*

The term *ecology* is used here to emphasize interdependence rather than to imply a single causal sequence. Feedback does not operate as an isolated instructional medium; it is produced and experienced through relationships among sources, delivery modes, learners, teachers, tasks, purposes, technological affordances, and institutional constraints. A change in one element can therefore alter the function or usability of another. For example, the explanatory richness of audio may be advantageous when learners can process spoken feedback efficiently, but less useful when listening confidence is low; similarly, a multimodal configuration may increase informational richness while also increasing information load or technical demands.

The ecological perspective also distinguishes feedback resources from their use. A mode may possess particular affordances without those affordances necessarily being taken up by learners or translated into improved writing. The reviewed evidence therefore suggests a sequence of related but non-equivalent processes: a feedback configuration creates particular affordances and constraints; these shape stakeholder experience; experience influences engagement and feedback use; and use may contribute to revision and writing outcomes. These relationships are neither uniform nor strictly linear, because contextual conditions can moderate each stage and evidence from use or outcomes may inform subsequent feedback design.

The Complementary Feedback Ecology Model extends this perspective by integrating three dimensions that are often examined separately: (a) feedback source and delivery mode, (b) contextual conditions that shape their usability and pedagogical fit, and (c) the relationship between stakeholder experience and actual feedback use. Its contribution is therefore not to propose that one feedback configuration causes a particular outcome, but to provide a framework for explaining why apparently divergent findings can coexist across studies. The model is heuristic and hypothesis-generating rather than a validated causal model; its value lies in specifying relationships that can be tested through future research.

4.6 *Contribution to the Literature*

The review's central contribution is conceptual and integrative rather than simply descriptive. It moves the field beyond feedback-mode competition by distinguishing feedback source from delivery mode and by integrating evidence across stakeholder experience, contextual conditions, feedback use, and writing-related outcomes. This synthesis shows why positive perceptions cannot be assumed to indicate uptake or learning and why no source or mode can be judged independently of its instructional configuration. The Complementary Feedback Ecology Model formalizes this contribution as a heuristic framework for organizing these relationships and generating testable propositions.

5. IMPLICATIONS

5.1 *Implications for Teachers*

Feedback modes should be selected according to pedagogical function rather than technological novelty. Written feedback is useful when precision and retrievability are primary; audio when elaboration and interpersonal support are important; audiovisual feedback when explanation and visual localization need to operate together; and AI when rapid preliminary analysis or additional feedback is useful, followed by teacher verification where contextual or rhetorical judgment matters.

5.2 *Implications for Students*

Students need support in using feedback, not merely receiving it. Instruction should help learners locate and interpret feedback, evaluate AI-generated recommendations, decide which suggestions to accept or reject, and convert feedback into concrete revision actions.

5.3 *Implications for AI Implementation*

Institutions should avoid treating AI output as a substitute for teacher judgment. A defensible implementation sequence is AI analysis → teacher verification → contextualization → student feedback → student evaluation and revision. This sequence preserves potential gains in speed and coverage while maintaining human oversight.

6. RESEARCH GAPS AND FUTURE DIRECTIONS

6.1 *Longitudinal and Process-Oriented Research*

Future studies should examine feedback across repeated writing cycles and determine whether perceptions, preferences, and use change with experience. Combining perception measures with feedback logs, revision analysis, and delayed writing measures would provide stronger evidence about the pathway from feedback experience to learning.

6.2 *Direct Student–Teacher Comparisons*

Claims of stakeholder convergence should be based on designs that examine students and teachers under comparable conditions. Future research should directly compare their evaluations of the same feedback configurations rather than relying primarily on cross-study inference.

6.3 *Workload-Sensitive Designs*

Studies should report preparation time, class size, feedback length, number of feedback points, technical preparation, and teacher cognitive load. Such measures would clarify when richer feedback is pedagogically valuable and institutionally sustainable.

6.4 *Better-Specified AI Comparisons*

AI research should report model version, prompt, feedback criteria, writing genre, proficiency level, feedback sequence, and degree of teacher feedback procedure. Greater specification is necessary to explain heterogeneous findings and permit meaningful replication and comparison.

6.5 *Integrated Measurement of Perception, Use, and Learning*

Future studies should combine perceptions with evidence of actual feedback use, revision, and longer-term writing development. This would help determine when perceived usefulness translates into observable uptake and learning.

7. LIMITATIONS

Several limitations qualify the synthesis. First, this is a critical integrative review rather than a registered systematic review. Although the evidence-identification process was structured and verification-oriented, it does not permit reconstruction of an exhaustive database-level search; therefore, the review makes no claim of complete literature coverage. Second, the evidence base is heterogeneous in participants, settings, tasks, technologies, intervention lengths, and outcome measures, which limits direct comparison across studies and precludes the calculation of a pooled effect size. Third, the geographical distribution of the literature is uneven, with substantial representation from particular higher-education and EFL contexts. This may limit the transferability of the synthesis to other educational and linguistic settings. Finally, most student–teacher convergence claims are interpretive because relatively few studies examine both groups under comparable conditions. The synthesis should therefore be understood as hypothesis-generating, particularly in relation to the proposed model and the rapidly evolving AI evidence.

8. RECOMMENDATIONS

The limitations suggest that future research should focus on analyses that compare student and teacher perspectives simultaneously rather than inferring consensus based on separate surveys. The future research should also analyze perceptions and actual feedback, with the revision of long-term writing progress, to see if the requested practicality improves writing skills.

Due to the rapid advancements of AI technologies, researchers should disclose the AI models they employed, as well as the prompts and feedback criteria; the genre of writing, the level of students, and the degree of teacher interference. Further studies should determine and regulate the relationships among feedback source, feedback context, students' perceptions, feedback practices, and the final writing products in different educational settings. These recommendations should improve the construct validity of feedback research in higher education L2 writing.

9. CONCLUSION

This critical integrative review suggests that feedback in higher-education L2 writing is better examined through the relationships among feedback source, delivery mode, context, stakeholder experience, and feedback use than through a universal ranking of feedback modes. Written, audio, and audiovisual feedback offer different affordances, while teacher, AI, and human–AI

configurations represent different feedback sources and processes. Across these configurations, perceived value does not necessarily correspond to uptake, revision, or learning outcomes.

The review's main contribution is to frame feedback as an ecology of interacting sources, modes, and contextual conditions rather than as a competition among delivery modes. Distinguishing feedback source from delivery mode also provides a clearer basis for interpreting findings across studies. The Complementary Feedback Ecology Model synthesizes these relationships as a heuristic framework; it is not intended to represent a validated causal model.

Future research should prioritize direct comparisons of feedback sources and modes, clearer reporting of contextual and procedural conditions, workload-sensitive designs, and closer examination of the relationship between perceptions, feedback use, and learning outcomes. These priorities may help clarify when and how different feedback configurations are most useful in higher-education writing, particularly as AI-mediated feedback continues to develop.

REFERENCES

Abdu, Murad, and Saeed Mohammed. 2021. "Does Teacher Feedback Mode Matter for Language Students?" *Asian EFL Journal* 28 (1): 202–28. https://www.asian-efl-journal.com/monthly-editions-new/2021-monthly-edition/volume-28-issue-1-1-february-2021/index.htm.

Alharbi, Mohammed Abdullah. 2022. "Exploring the Impact of Teacher Feedback Modes and Features on Students' Text Revisions in Writing." *Assessing Writing* 52 (April): 100610. https://doi.org/10.1016/j.asw.2022.100610.

Alharbi, Mohammed Abdullah, and Abdurrazzag Alghammas. 2021. "Teacher Written Vs. Audio Feedback on Undergraduates' Written Assignments." *Theory and Practice in Language Studies* 11 (12): 1562–70. https://doi.org/10.17507/tpls.1112.08.

Ali, Amira Desouky. 2016. "Effectiveness of Using Screencast Feedback on EFL Students' Writing and Perception." *English Language Teaching* 9 (8): 106. https://doi.org/10.5539/elt.v9n8p106.

Alnemrat, Areen, Hesham Aldamen, Mohamad Almashour, Mutasim Al-Deaibes, and Rami AlSharefeen. 2025. "AI Vs. Teacher Feedback on EFL Argumentative Writing: A Quantitative Study." *Frontiers in Education* 10 (July): 1614673. https://doi.org/10.3389/feduc.2025.1614673.

Bakla, Arif. 2020. "A Mixed-Methods Study of Feedback Modes in EFL Writing." *Language Learning & Technology* 24 (1): 107–28. https://doi.org/10.64152/10125/44712.

Bezemer, Jeff, and Gunther Kress. 2008. "Writing in Multimodal Texts: A Social Semiotic Account of Designs for Learning." *Written Communication* 25 (2): 166–95. https://doi.org/10.1177/0741088307313177.

Brearley, Francis Q., and W. Rod Cullen. 2012. "Providing Students With Formative Audio Feedback." *Bioscience Education* 20 (1): 22–36. https://doi.org/10.11120/beej.2012.20000022.

Bush, Jerome C. 2020. "Using Screencasting to Give Feedback for Academic Writing." *Innovation in Language Learning and Teaching* 15 (5): 1–14. https://doi.org/10.1080/17501229.2020.1840571.

Cao, Lien. 2023. "EFL Undergraduates' Perceptions of the Use of Screencast Feedback in an Online Writing Course." *TESOL and Technology Studies* 4 (2): 1–14. https://doi.org/10.48185/tts.v4i2.786.

Cavaleri, Michelle, Satomi Kawaguchi, Bruno Di Biase, and Clare Power. 2019. "How Recorded Audio-Visual Feedback Can Improve Academic Language Support." *Journal of University Teaching and Learning Practice* 16 (4). https://doi.org/10.53761/1.16.4.6.

Cavanaugh, Andrew, and Liyan Song. 2014. "Audio Feedback Versus Written Feedback: Instructors' and Students' Perspectives." *MERLOT Journal of Online Learning and Teaching* 10 (1): 122–38. https://jolt.merlot.org/vol10no1/cavanaugh_0314.pdf.

Cronin, Matthew A., and Elizabeth George. 2020. "The Why and How of the Integrative Review." *Organizational Research Methods* 26 (1): 168–92. https://doi.org/10.1177/1094428120935507.

Cunningham, Kelly J. 2019. "Student Perceptions and Use of Technology-Mediated Text and Screencast Feedback in ESL Writing." *Computers and Composition* 52 (June): 222–41. https://doi.org/10.1016/j.compcom.2019.02.003.

ElEbyary, Khaled, Ramy Shabara, and Deena Boraie. 2024. "The Differential Role of AI-Operated WCF

in L2 Students' Noticing of Errors and Its Impact on Writing Scores." *Language Testing in Asia* 14 (1). https://doi.org/10.1186/s40468-024-00312-1.

Han, Feifei, and Zehua Wang. 2026. "Teacher Feedback and ChatGPT Feedback on Chinese University EFL Learners' English Essay Revision: A Mixed-Methods Study." *Computers and Education: Artificial Intelligence* 10 (June): 100590. https://doi.org/10.1016/j.caeai.2026.100590.

Han, Jining, and Mimi Li. 2024. "Exploring ChatGPT-Supported Teacher Feedback in the EFL Context." *System* 126 (November): 103502. https://doi.org/10.1016/j.system.2024.103502.

Karagoz, Inanc. 2025. "AI-Generated Feedback in English Writing Instruction for Language Learners: A Systematic Review." *The Reading Matrix: An International Online Journal* 25 (1): 51–67. https://www.readingmatrix.com/files/34-9gp90gjo.pdf.

Keane, Kjrsten, Daniel McCrea, and Miriam Russell. 2018. "Personalizing Feedback Using Voice Comments." *Open Praxis* 10 (4): 309. https://doi.org/10.5944/openpraxis.10.4.909.

Khojasteh, Laleh, Afshin Soori, and Fareeha Javed. 2025. "Comparing Teacher E-Feedback, AI Feedback, and Hybrid Feedback in Enhancing EFL Writing Skills." *Technology in Language Teaching & Learning* 7 (3): 102626. https://doi.org/10.29140/tltl.v7n3.102626.

Kress, Gunther R. 2001. *Multimodal Discourse : The Modes and Media of Contemporary Communication* . Arnold. https://search.worldcat.org/title/Multimodal-discourse-:-the-modes-and-media-of-contemporary-communication/oclc/882785470.

Liu, Chenying, and Maocheng Liang. 2026. "The Effects of AI-Assisted Feedback on Students' Perceptions, Emotions, and Learning Outcomes in L2 Writing: A Three-Level Meta-Analysis." *System* 139 (June): 104018. https://doi.org/10.1016/j.system.2026.104018.

Lunt, Tom, and John Curran. 2009. "'Are You Listening Please?' The Advantages of Electronic Audio Feedback Compared to Written Feedback." *Assessment & Evaluation in Higher Education* 35 (7): 759–69. https://doi.org/10.1080/02602930902977772.

Matikainen, Tiina. 2024. "Academic Writing in English: Lessons From an EMI-Program in Japan." *Journal of English for Academic Purposes* 68 (March): 101358. https://doi.org/10.1016/j.jeap.2024.101358.

Nazaretsky, Tanya, Paola Mejia-Domenzain, Vinitra Swamy, Jibril Frej, and Tanja Käser. 2024. "AI or Human? Evaluating Student Feedback Perceptions in Higher Education." *Technology Enhanced Learning for Inclusive and Equitable Quality Education: 19th European Conference on Technology Enhanced Learning, EC-TEL 2024,* (Krems, Austria), 284–98.

Nemec, Eric C., and Matthew Dintzner. 2016. "Comparison of Audio Versus Written Feedback on Writing Assignments." *Currents in Pharmacy Teaching and Learning* 8 (2): 155–59. https://doi.org/10.1016/j.cptl.2015.12.009.

Oanh, Tran Thi Thuy, and Tran Thi Minh Ngoc. 2026. "Teachers' and Students' Perceptions of AI-Based Writing Assessment." *The University of Danang - Journal of Science and Technology* 24 (2): 142–48. https://doi.org/10.31130/ud-jst.2026.24(2).006e.

Page, Matthew J, Joanne E McKenzie, Patrick M Bossuyt, et al. 2021. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *British Medical Journal* 372 (71). https://doi.org/10.1136/bmj.n71.

Rethlefsen, Melissa L., Shona Kirtley, Siw Waffenschmidt, et al. 2021. "PRISMA-S: An Extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews." *Systematic Reviews* 10 (1). https://doi.org/10.1186/s13643-020-01542-z.

Roy, Sukanto, and Matthew A. Vetter. 2023. "Audio Versus Written Feedback: Multilingual Students' Perspective in a U.S. Composition Class." *Feedback Research in Second Language* 1 (December): 149–74. https://doi.org/10.32038/frsl.2023.01.09.

Ryan, Tracii, Michael Henderson, and Michael Phillips. 2019. "Feedback Modes Matter: Comparing Student Perceptions of Digital and Non-Digital Feedback Modes in Higher Education." *British Journal of Educational Technology* 50 (3): 1507–23. https://doi.org/10.1111/bjet.12749.

Saeed, Murad Abdu, Atef AbuSa'aleek, and Mohammed Abdullah Alharbi. 2024. "Examining Teacher's Evaluative Language in Written, Audio and Screencast Feedback on EFL Learners' Writing From the Appraisal Framework: A Linguistic Perspective." *Assessing Writing* 61 (July): 100871. https://doi.org/10.1016/j.asw.2024.100871.

Sasaki, Akihiko, and Osamu Takeuchi. 2026. "'Noticed but Not Used': The Gap Between AI Feedback and Its Delayed Application in L2 Academic Writing." *Asian Journal of English Language Teaching* 35 (4): 80–100. https://doi.org/10.65961/ajelt-2026-4-005.

Torraco, Richard J. 2005. "Writing Integrative Literature Reviews: Guidelines and Examples." *Human Resource Development Review* 4 (3): 356–67. https://doi.org/10.1177/1534484305278283.

Voelkel, Susanne, and Luciane V. Mello. 2014. "Audio Feedback – Better Feedback?" *Bioscience Education* 22 (1): 16–30. https://doi.org/10.11120/beej.2014.00022.

Zeevy-Solovey, Orit. 2024. "Comparing Peer, ChatGPT, and Teacher Corrective Feedback in EFL Writing: Students' Perceptions and Preferences." *Technology in Language Teaching & Learning* 6 (3): 1482. https://doi.org/10.29140/tltl.v6n3.1482.

Özkul, Sertaç, and Deniz Ortaçtepe. 2017. "The Use of Video Feedback in Teaching Process-Approach EFL Writing." *TESOL Journal* 8 (4): 862–77. https://doi.org/10.1002/tesj.362.