

INPLASY

Hybrid Human–Generative AI Feedback in Formal Reciprocal Peer Observation Programs in Higher Education: A Systematic Review and Potential Meta-Analysis

INPLASY202680039

doi: 10.37766/inplasy2026.8.0039

Received: 10 August 2026

Published: 10 August 2026

Cuevas-Aburto, JA; Fuentes-Rifo, KA; Tamarín-Tamarín, NA.

Corresponding author:

Jorge Cuevas Aburto

jorgecuevas@santotomas.cl

Author Affiliation:

Learning Center, National Undergraduate Studies Office, Vice-Rectoría for Academic Affairs, Research and Postgraduate Studies, Universidad Santo Tomás, campus Los Ángeles, Chile; Doctoral Program in Education, Universidad de Concepción, Concepción, Chile.

ADMINISTRATIVE INFORMATION

Support - No external funding is declared. The study is conducted using the research team's own resources within the scope of their institutional academic duties.

Review Stage at time of this submission - The review has not yet started.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202680039

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 10 August 2026 and was last updated on 14 August 2026.

INTRODUCTION

Review question / Objective Among in-service university faculty participating in formal Reciprocal Peer Observation programs (P), do hybrid human–generative AI feedback systems (I), compared with exclusively human feedback, exclusively automated feedback, or the absence of a formal program (C), produce improvements in the dialogic quality of feedback (primary outcome) and reductions in evaluator time burden (secondary outcome) (O), in studies published between January 1, 2018, and August 17, 2026 (T), and to what extent are these effects moderated by sociogeographic context, particularly in Latin America (M)?

Objectives:

1. To synthesize available evidence on the effect of hybrid human–generative AI systems on the dialogic quality of feedback within formal Reciprocal Peer Observation programs in higher education.

2. To estimate quantitatively the effect of these systems on the evaluator's temporal efficiency.
3. To examine whether sociogeographic context, with an emphasis on Latin America, moderates the effects identified in the preceding objectives.

Rationale Reciprocal peer observation (RPO) is a core component of formal pedagogical support programs in higher education because, given optimal conditions such as structured design, institutional support, flexible leadership, and training, it remains an effective modality for generating sustained change in pedagogical practice (Johnston et al., 2022; Zeng, 2020). However, most interventions omit elements indispensable for a high-quality feedback cycle due to time constraints, lack of conceptual scaffolding, and a tendency toward generic comments (Flores et al., 2025; Kerman et al., 2023; Qi et al., 2026; Ribosa et al., 2024). Consequently, scaling high-quality RPO exclusively through human effort is structurally unfeasible (Maheshi et

al., 2024). To address this limitation, hybrid human-generative AI systems offer a scalable, personalizable sociotechnical response (Kaliisa et al., 2026), provided they are properly calibrated (C.-Q. Huang et al., 2026; Gu & Yan, 2025; Y. Huang et al., 2026). Despite this potential, empirical evidence reveals a critical gap; existing literature predominantly focuses on students, systematically excluding in-service university faculty as feedback recipients or providers (Bond, 2024; Nedrehagen et al., 2025). Furthermore, pedagogical support programs in Latin America operate under heterogeneous digital infrastructures and weak regulatory frameworks compared to the Global North (UNESCO-IESALC, 2025), and are severely underrepresented in the literature (Bond, 2024; Crompton & Burke, 2023). Therefore, broadening the search horizon to include regional databases is required to counteract indexing bias and capture these qualitatively distinct realities (Hartuique et al., 2025).

Condition being studied Condition being studied: This review examines the dialogic quality and temporal efficiency of faculty-to-faculty feedback exchanged within formal pedagogical support programs in higher education, specifically where such feedback is produced, supported, or mediated by hybrid human and generative AI systems. In this context, formal programs are operationalized as institutionally documented structures comprising one or more complete feedback cycles. Each cycle must include teaching observation or review followed by a systematized and verifiable feedback mechanism. This definition allows single-cycle studies to be included when they provide sufficient documentation of the observation or review process and its associated feedback procedure, while still excluding informal, spontaneous, or undocumented exchanges. The number of complete cycles reported in each study, $k=1$ versus $k \geq 2$, will be extracted as a data item and examined as a potential moderating variable in subgroup and sensitivity analyses, rather than being used as an eligibility threshold. This approach preserves the sensitivity of the review, reduces the risk of excluding informative studies, and allows variation in program intensity to be analysed empirically.

Eligible modalities include reciprocal peer observation, peer review of teaching, institutional pedagogical coaching, and faculty support structures with an explicit evaluative component. The condition of interest is twofold. First, the review examines the dialogic quality of peer feedback, operationalized across five subdimensions, specificity, personalization, knowledge-building markers, orientation toward

self-regulation, and dialogic tone. These dimensions address the documented tendency of unsupported human feedback to remain generic, weakly personalized, or insufficiently dialogic. Second, the review examines the temporal efficiency of faculty evaluators, including time per cycle, number of iterations, and perceived workload. This dimension responds to the structural time and scaffolding constraints that limit the scalability of high-quality feedback when it depends exclusively on human evaluators.

Both dimensions are examined in relation to in-service university faculty acting as feedback providers, feedback recipients, or both. This focus is necessary because the existing literature on generative AI-mediated feedback has concentrated mainly on students, leaving faculty-to-faculty pedagogical feedback comparatively underexamined. Accordingly, the review centres on formal institutional feedback practices in higher education and assesses whether hybrid human and generative AI systems can improve feedback quality, reduce evaluator workload, or reshape the dialogic conditions under which university teachers engage in pedagogical improvement.

METHODS

Search strategy Search strategy: The search strategy is organised into three conceptual blocks combined with AND. The first block captures generative AI systems, large language models, and hybrid feedback tools, including "AI-mediated feedback", "generative AI-assisted formative assessment", "dialogic AI feedback", and "automated feedback system" when the interface allows extended Boolean syntax. The second block captures faculty mentoring, peer observation, pedagogical coaching, and teaching feedback. The third block limits retrieval to higher education faculty and related university teaching populations. Within each block, synonymous terms are connected with OR to maintain sensitivity across English- and Spanish-language literature. Database-specific syntaxes were adapted to the field codes, controlled vocabularies, parsing rules, and Boolean limitations of each source. The search period covers January 1, 2018, to the effective closing date of the search in each database, not later than the close of the seven-calendar-day window beginning August 10, 2026, with dates recorded per source in the search log.

For Scopus, the search will be conducted in titles, abstracts, and keywords using the following base syntax, TITLE-ABS-KEY (("hybrid feedback" OR "human-AI feedback" OR "AI-assisted feedback" OR "AI-mediated feedback" OR "generative AI" OR "GenAI" OR "generative AI-assisted formative

assessment" OR "dialogic AI feedback" OR "automated feedback system*" OR "large language model*" OR LLM OR "ChatGPT" OR "GPT-4" OR "artificial intelligence") AND ("peer observation" OR "peer review of teaching" OR "peer coaching" OR "classroom observation" OR "pedagogical coaching" OR "instructional coaching" OR "teaching mentoring" OR "acompañamiento pedagógico" OR "observación entre pares" OR "retroalimentación docente") AND ("higher education" OR "university" OR "tertiary education" OR "educación superior") AND ("teacher*" OR "faculty" OR "instructor*" OR "lecturer*" OR "professor*" OR "docente*")) AND PUBYEAR > 2017 AND PUBYEAR < 2027.

For Web of Science Core Collection, focusing on the Social Sciences Citation Index, SSCI, and Emerging Sources Citation Index, ESCI, the Topic field, TS, will be used across titles, abstracts, and keywords. Unsupported proximity operators are omitted. Regional and grey literature indexes, such as SciELO Citation Index and ProQuest Dissertations & Theses Citation Index, will be excluded in this interface to avoid duplication with database-specific searches. The publication window, 2018 to 2026, will be applied through the platform's publication year filter rather than embedded in the Topic field. The syntax is TS=(("hybrid feedback" OR "human-AI feedback" OR "AI-assisted feedback" OR "AI-mediated feedback" OR "generative AI" OR "GenAI" OR "generative AI-assisted formative assessment" OR "dialogic AI feedback" OR "automated feedback system*" OR "large language model*" OR LLM OR "ChatGPT" OR "GPT-4" OR "artificial intelligence") AND ("peer observation" OR "peer review of teaching" OR "peer coaching" OR "classroom observation" OR "pedagogical coaching" OR "instructional coaching" OR "teaching mentoring" OR "acompañamiento pedagógico" OR "observación entre pares" OR "retroalimentación docente") AND ("higher education" OR "university" OR "tertiary education" OR "educación superior") AND ("teacher*" OR "faculty" OR "instructor*" OR "lecturer*" OR "professor*" OR "docente*")).

For ERIC, via EBSCO, the syntax combines ERIC thesaurus descriptors with free-text terms to capture both indexed literature and emerging studies not yet consistently assigned controlled descriptors, (("hybrid feedback" OR "human-AI feedback" OR "AI-assisted feedback" OR "AI-mediated feedback" OR "generative AI" OR "generative AI-assisted formative assessment" OR "dialogic AI feedback" OR "automated feedback system" OR "large language model" OR "ChatGPT" OR "GPT-4" OR "artificial intelligence") AND (DE "Faculty Development" OR DE "Peer

Evaluation" OR DE "Teacher Evaluation" OR TX "peer observation" OR TX "pedagogical coaching" OR TX "instructional coaching") AND (DE "Higher Education" OR DE "Colleges" OR DE "Universities") AND (DE "Faculty" OR DE "College Faculty" OR TX "university teacher*" OR TX "university instructor*" OR TX "university lecturer*")) AND PY 2018-2026. The combined use of DE descriptors and TX free-text terms is retained to reduce irrelevant retrieval while preserving recently emerging terminology related to generative AI.

For PsycINFO, via APA PsycNET, the syntax uses descriptors from the APA Thesaurus of Psychological Index Terms and combines them with title and abstract searches, (DE "Artificial Intelligence" OR TI "hybrid feedback" OR TI "human-AI feedback" OR TI "AI-assisted feedback" OR TI "AI-mediated feedback" OR TI "generative AI" OR TI "generative AI-assisted formative assessment" OR TI "dialogic AI feedback" OR TI "automated feedback system*" OR AB "large language model*" OR TI "ChatGPT" OR TI "GPT-4") AND (DE "Faculty Development" OR DE "Peer Evaluation" OR TI "peer observation" OR AB "pedagogical coaching" OR AB "instructional coaching") AND (DE "Higher Education" OR DE "Colleges and Universities") AND (DE "Faculty (Academic)" OR TI "university teacher*" OR AB "university instructor*") AND PY 2018-2026.

For Education Source, via EBSCO, the search will use the database's controlled descriptors together with free-text terminology, (DE "Artificial intelligence" OR TX "generative AI") AND (DE "Peer evaluation" OR TX "peer observation") AND (DE "Higher education" OR DE "Universities and colleges").

For the ACM Digital Library and IEEE Xplore, a simplified metadata-level syntax will be applied because both platforms are sensitive to long nested Boolean strings, ("generative AI" OR "large language model") AND ("peer observation" OR "teacher feedback") AND ("higher education").

For SciELO, the search will be decomposed into three conceptual modules because the native interface is vulnerable to parsing failures when processing long, nested Boolean expressions. The modules correspond to technological intervention, pedagogical context, and higher education population. Searches will be run sequentially in title fields and then in abstract fields to improve coverage while preserving search stability. Search Box 1, Field, Título / Resumen, "inteligencia artificial" OR "IA generativa" OR "GenAI" OR "LLM" OR "retroalimentación híbrida" OR "feedback híbrido". AND Search Box 2, Field, Título / Resumen, "acompañamiento pedagógico" OR "observación entre pares" OR "observación de

aula" OR "retroalimentación docente" OR "coaching pedagógico". AND Search Box 3, Field, Título / Resumen, "educación superior" OR "universidad". No thematic filters will be applied, in order to avoid excluding interdisciplinary studies. Results will be restricted by publication year, 2018 to 2026, and language, Spanish and English. Conceptual eligibility will be assessed later through double-blind screening in Rayyan.

For Redalyc, the search will use title and abstract fields because the interface does not support complex Boolean searches with controlled descriptors, (titulo:("inteligencia artificial" OR "IA generativa" OR "retroalimentación") AND titulo:("acompañamiento pedagógico" OR "observación entre pares" OR "docente universitario")) OR (resumen:("inteligencia artificial" OR "IA generativa") AND resumen:("feedback" OR "retroalimentación") AND resumen:("educación superior" OR "universidad")). The filter año:[2018 TO 2026] will be applied after retrieval.

For Dialnet, the strategy will use two modular free-text queries to capture Spanish-language literature on generative AI, pedagogical observation, and hybrid or AI-mediated feedback in university teaching, while avoiding the platform's limitations with long nested Boolean strings. Module 1, Generative Technology and Pedagogical Observation, ("IA generativa" OR "modelos de lenguaje") AND ("feedback docente" OR "acompañamiento pedagógico" OR "observación de aula" OR "observación entre pares") AND universidad. Module 2, Hybrid and AI-Mediated Feedback, ("feedback híbrido" OR "retroalimentación híbrida" OR "feedback mediado por IA") AND universidad AND docente. Both modules will be run as free-text searches across available titles, abstracts, keywords, and descriptive metadata. Retrieval will be constrained by the predefined temporal window and eligible languages, with all subsequent judgments regarding conceptual relevance and study eligibility reserved for the double-blind screening stage conducted in Rayyan.

For ProQuest Dissertations & Theses Global, the strategy will use the noft, Anywhere except full text, field code to capture graduate-level grey literature while limiting false positives from full-text dissertation searches, noft("generative artificial intelligence" OR "generative AI" OR "large language model*" OR LLM* OR "human-AI hybrid" OR "hybrid feedback" OR "AI-mediated feedback" OR "generative AI-assisted" OR "dialogic AI feedback" OR "automated feedback system*" OR "IA generativa" OR "modelos de lenguaje" OR "feedback híbrido" OR "retroalimentación híbrida" OR "feedback mediado por IA") AND noft("peer observation" OR "reciprocal peer observation" OR

"peer review of teaching" OR "pedagogical support" OR "faculty support" OR "teaching feedback" OR "pedagogical feedback" OR "observación de aula" OR "observación entre pares" OR "acompañamiento pedagógico" OR "feedback docente" OR "retroalimentación docente") AND noft("higher education" OR universit* OR faculty OR "in-service teacher*" OR "university teacher*" OR "university instructor*" OR "educación superior" OR "profesorado universitario" OR "docente universitario"). Results will be restricted to the 2018 to 2026 temporal scope using the platform's date filters.

For EdArXiv, accessed through OSF Preprints, the search will use short modular queries because the interface has limited support for nested Boolean logic. The following searches will be applied, "generative AI" feedback university; "hybrid feedback" OR "AI-mediated feedback" OR "generative AI-assisted"; and "feedback híbrido" OR "retroalimentación híbrida". Searches will be filtered by Provider, EdArXiv, and Resource Type, Preprint. The year range restriction will be applied manually during screening in Rayyan, since OSF date filtering may be unreliable for custom ranges.

For SSRN, the search will use the Advanced Search field Title, Abstract & Keywords. The syntax is intentionally simplified to avoid platform failures with long nested strings while retaining the essential terms, ("generative AI" OR "GenAI" OR "LLM" OR "large language model") AND ("peer observation" OR "peer coaching" OR "classroom observation" OR "teacher feedback") AND ("higher education" OR "university"). No network or sub-network filters will be applied, since restricting the search to a specific SSRN network could exclude relevant interdisciplinary records. Because custom date limits may not apply consistently in SSRN, all results will be exported and the 2018 to 2026 restriction will be enforced during double-blind screening in Rayyan.

For Google Scholar, the search will use a streamlined Boolean string because the platform relies on broad full-text indexing, has limited field-specific operators, and is sensitive to long queries, ("hybrid feedback" OR "generative AI" OR "GenAI" OR "LLM" OR "large language model") AND ("peer observation" OR "peer coaching" OR "classroom observation" OR "teacher feedback") AND ("higher education" OR "university"). The publication window, 2018 to 2026, will be applied using Google Scholar's custom date-range filter. Screening will focus on the first 200 records ranked by relevance, with access dates documented. Final eligibility decisions will be made during double-blind screening in Rayyan.

Participant or population Participant or population: The broader population of interest comprises in-service university faculty working in higher education institutions. Within this population, eligible participants are defined as faculty members who actively engage as peer evaluators, feedback providers, and/or observees, feedback recipients, in formal pedagogical support programs. For the purposes of this review, these programs are operationalized as institutional initiatives with a documented structure that includes at least one complete cycle of teaching observation or review, followed by a systematized and verifiable faculty-to-faculty feedback mechanism. This criterion is intended to ensure that included studies examine structured pedagogical feedback processes, while avoiding the unnecessary exclusion of relevant evidence from programs that report a single documented cycle. Accordingly, informal, spontaneous, or undocumented exchanges will be excluded, as they do not provide sufficient methodological traceability for analysis. Eligible modalities include reciprocal peer observation, peer review of teaching, institutional pedagogical coaching, and faculty support structures with an explicit evaluative feedback component. Rather than being used as an eligibility threshold, the number of complete cycles documented in each study will be extracted as a data item and explored as a potential moderator of effect size. This approach preserves the sensitivity of the review, reduces the risk of excluding informative studies, and allows variation in program intensity to be examined analytically. Studies centered on students, preschool, primary or secondary educators, pre-service teachers, or professional development programs lacking a formal feedback component will be excluded.

Intervention Intervention The eligible intervention is any hybrid human-generative AI system implemented within a formal program of observation of teaching practice in higher education. Eligible interventions encompass any hybrid workflow combining human agency and Generative AI in feedback production. Generative AI (GenAI) is based on Large Language Models (LLMs) or other architectures grounded in transformer mechanisms and their associated attention mechanisms, designed to infer and produce novel natural-language text (OECD, 2024; UNESCO, 2023). Four eligible configurations are considered as a descriptive taxonomy, not as a closed exclusion barrier: (a) draft-review, in which AI generates an initial feedback draft that the human evaluator reviews and adjusts; (b) detection-elaboration, in which AI identifies

markers in the observed practice that the instructor elaborates into feedback; (c) parallel-synthesis, in which AI and human generate independent feedback subsequently integrated; and (d) real-time assistance, in which AI provides live, adaptive scaffolding during observation. Studies describing other documented hybrid workflows are also eligible when they show that human agency and Generative AI jointly contribute to feedback production. Systems combining generative and non-generative components are eligible only where the generative component effectively produces or substantively supports the textual feedback delivered to the instructor; this hybrid feature is coded as an intervention variable for heterogeneity exploration. Studies must explicitly document the GenAI technology, model, and, where available, version used.

Comparator Three valid comparators are defined: (1) faculty-to-faculty feedback produced exclusively by human evaluators within a formal pedagogical support program; (2) feedback generated exclusively by a GenAI system without human intervention in content elaboration, though non-editorial oversight (e.g., selecting which observation is submitted) may exist; and (3) absence of a formal support program for in-service teachers in higher education. Studies comparing the hybrid system with non-generative AI systems (e.g., rule-based automated scoring) are recorded separately as exploratory secondary comparators and not combined with pure-GenAI comparisons. A single study may include more than one comparator, each analyzed independently respecting the original unit of randomization or comparison.

Study designs to be included Eligible designs include randomized controlled trials, quasi-experimental studies with an equivalent control group or statistical adjustment, longitudinal observational studies with between-group comparison, mixed-methods studies with an evaluable primary quantitative component, and rigorous qualitative studies. Single-case designs will be analyzed separately as exploratory evidence. Prior systematic reviews will be utilized exclusively for theoretical contrast and backward citation searching, not as primary units of analysis.

Eligibility criteria Eligibility criteria: Included: studies examining in-service university faculty within formal pedagogical support programs comprising one or more documented feedback cycles, evaluating hybrid human-GenAI feedback systems against exclusively human feedback, exclusively GenAI feedback, or absence of a

formal program, on dialogic feedback quality and/or evaluator temporal efficiency, published in Spanish or English between January 1, 2018, and the search closure date. Single-cycle studies are eligible where at least one complete cycle of observation or review of teaching performance followed by a systematized, verifiable feedback mechanism is documented. The number of complete cycles reported in each study, $k=1$ versus $k \geq 2$, will be extracted as a data item and explored as a potential moderating variable in subgroup and sensitivity analyses, rather than applied as an eligibility threshold. Eligible designs: RCTs, quasi-experimental studies, longitudinal observational comparisons, mixed-methods studies, and qualitative studies with explicit rigor criteria; single-case designs analyzed separately. Excluded: studies centered on students, non-university-level or pre-service teachers, professional development without a formal feedback component; interventions using non-generative AI as the sole comparator without a qualifying primary comparator; studies not reporting disaggregated data for in-service university faculty when populations are mixed; studies in languages other than Spanish or English.

Information sources Searches will be executed in Scopus (Elsevier), Web of Science Core Collection with emphasis on SSCI and ESCI (Clarivate), ERIC via EBSCO, PsycINFO via APA PsycNET, Education Source (EBSCO), ACM Digital Library, IEEE Xplore, SciELO, Redalyc, Dialnet, and ProQuest Dissertations & Theses Global for graduate-level gray literature. To mitigate publication bias, discipline-specific preprint repositories EdArXiv and SSRN will also be searched. An additional Google Scholar search will be conducted, limited to the first 200 records ranked by relevance per thematic block, with access date documented, acknowledging that platform's reproducibility limitations. Regional Latin American databases (SciELO, Redalyc, Dialnet) are treated as primary, not complementary, sources given documented Latin American underrepresentation in Anglophone indexing.

Main outcome(s) The primary outcome is dialogic quality of feedback, operationalized across five subdimensions: specificity of content, personalization to the observed practice, knowledge-building markers (adapted exploratorily from Scardamalia & Bereiter, 2005), orientation to self-regulation, and dialogic tone. In the absence of a standardized instrument for this construct in AI-mediated peer observation, these dimensions will be operationalized through a low-inference qualitative codebook based on directed qualitative

content analysis (Hsieh & Shannon, 2005), calibrated via Cohen's Kappa. This outcome will be synthesized primarily through narrative synthesis; meta-analysis is exploratory, contingent on standardized or transformable metrics and $k \geq 5$ studies.

Additional outcome(s) The secondary outcome is the faculty evaluator's temporal efficiency, measured continuously as time per feedback cycle, number of iterations, and perceived workload. Additional secondary outcomes, extracted when available, include the recipient's trust in the feedback, observable changes in pedagogical practice, faculty perceptions of agency and autonomy, and indicators of over-reliance on the generative AI system.

Data management Study selection follows independent double screening in two phases using Rayyan QCRI (web and mobile v. 1.7.7) with Blind Mode active until discrepancy resolution; access date and version are documented in the methodological log. Screening and discrepancy resolution run between August 17 and September 15, 2026. Inter-rater agreement is assessed via Cohen's Kappa (acceptance threshold $\kappa \geq 0.70$); persistent discrepancies are resolved by a third arbitrating reviewer. Data extraction uses a standardized form, piloted on five studies (September 11–15, 2026), by two independent reviewers extracting in duplicate, with discrepancies resolved by consensus. When data required for meta-analysis are missing, corresponding authors are contacted by email with a 21-day response window. For studies with multiple publications from the same sample, data are extracted once from the most complete publication. The complete process is reported via a PRISMA 2020 flow diagram.

Quality assessment / Risk of bias analysis Risk of bias is assessed by two independent reviewers using tools selected by study design: RoB 2 (Cochrane) for randomized controlled trials; ROBINS-I (Cochrane) for non-randomized studies with a comparator; CASP Qualitative Checklist for qualitative studies; MMAT (2018 version) for mixed-methods studies; and the Design and Evidence Standards for Single-Case Designs, What Works Clearinghouse (2022) Standards Handbook, for single-case designs. Prior systematic reviews are assessed with AMSTAR 2 solely for descriptive purposes regarding the state of the prior synthesis field; this does not constitute their inclusion as an analytic unit. Assessment uses Rayyan QCRI's Risk of Bias Tool in blinded mode, distinct from the screening module.

Strategy of data synthesis Structured narrative synthesis, following Popay et al. (2006), is the central integration method for the entire corpus, organized by outcome, comparator type, and hybrid configuration, using evidence summary tables. Meta-analysis proceeds only when $k \geq 5$ studies share a comparator, the outcome is standardizable or convertible to a common metric, and sufficient data are available; when $k < 5$, the corresponding objective is addressed narratively or, for Objective 3, qualitatively. A random-effects model is adopted given expected clinical and methodological heterogeneity; Hedges' g (continuous outcomes) and odds ratios (dichotomous outcomes) are calculated with 95% CIs. Heterogeneity variance (τ^2) is estimated via REML, with Hartung-Knapp-Sidik-Jonkman (HKSJ) adjustment for pooled-effect confidence intervals; DerSimonian-Laird is reported as sensitivity analysis. When $k \in \{5,6\}$, results are accompanied by prediction intervals. Analyses use R with the metafor package (Viechtbauer, 2010). Quantitative and qualitative evidence are integrated via joint displays (Guetterman et al., 2015), mapping effect sizes and temporal-efficiency metrics against analytic themes and Latin American structural conditions.

Subgroup analysis Predefined subgroup analyses examine sociogeographic context (Latin America versus other regions, per UNESCO-IESALC classification), hybrid configuration type (draft-review, detection-elaboration, parallel-synthesis, real-time assistance), comparator type, and implementation characteristics, conducted where study volume permits. When the Latin American subgroup yields $k < 5$, moderation is addressed through qualitative, descriptive, cross-study synthesis integrated within the joint displays rather than formal subgroup statistics.

Sensitivity analysis Sensitivity analyses will compare REML and Hartung-Knapp-Sidik-Jonkman estimates against conventional DerSimonian-Laird estimates. Further analyses will exclude studies assessed as having a high risk of bias and will evaluate results according to the availability of complete versus imputed data. When $k \in \{5, 6\}$, results will be interpreted with strict caution and accompanied by prediction intervals, given the high imprecision inherent in τ^2 estimation with small samples.

Language restriction The review includes exclusively Spanish and English studies. Portuguese and Mandarin are excluded due to linguistic constraints and database access barriers, constituting a formally declared coverage limitation.

Country(ies) involved Chile (review team). No geographic restrictions apply to the eligibility of primary studies; the country or region of origin will be extracted and incorporated as a moderating variable.

Other relevant information

Review not yet started.

Registration: INPLASY, registration number [INSERT INPLASY REGISTRATION NUMBER].

Review start date: 10-08-2026.

Anticipated completion date: 15-12-2026.

Any protocol amendment will be dated, justified and registered.

Protocol amendment (v2.0, 14 August 2026): Following a pilot screening phase, the eligibility criteria for Participants, specifically the minimum number of feedback cycles, and the search syntax technology and context blocks across databases were revised to prevent unjustified loss of relevant empirical evidence in this emerging field. The amendment does not alter the research question, objectives, comparators, or outcomes. Details of each modification are documented in the version history of this protocol.

Keywords Human-AI hybrid feedback; generative artificial intelligence; reciprocal peer observation; dialogic feedback; higher education; Latin America.

Dissemination plans The final manuscript will be submitted for open-access publication in an evidence-synthesis or education sciences journal. Findings will additionally be disseminated through presentations at academic congresses, scientific seminars, and specialized institutional lectures. Synthesis data and analytic materials will be deposited in an open data repository, such as the Open Science Framework, to ensure reproducibility. The complete protocol will remain available as supplementary material appended to the INPLASY registration.

Contributions of each author

Author 1 - Jorge Cuevas Aburto. - Corresponding author; led conception, design, and methodology; coordinated the review team; managed INPLASY registration; developed protocol, eligibility criteria, search, extraction, and synthesis plans; contributed to screening, full-text review, extraction, and verification; supervised risk-of-bias assessment, synthesis, and meta-analysis; drafted, revised, coordinated feedback, managed correspondence, and ensured integrity transparency, and process accountability.

Email: jorgecuevas@santotomas.cl

Author 2 - Karina Fuentes Rifo - Third arbitrating reviewer for unresolved discrepancies; methodological oversight and support in consensus decisions; doctoral supervisor of the corresponding author; academic guidance on the review design and manuscript development; support in author correspondence when required; critical manuscript revision; and approval of the final version.

Email: kafuente@udec.cl

Author 3 - Nicolás Tamarín - Contribution: Independent co-reviewer; title, abstract, and full-text screening using Rayyan QCRI; independent selection of eligible papers; data extraction and verification; risk-of-bias assessment; participation in discrepancy resolution; methodological and theoretical input on GenAI-mediated feedback, pedagogical support programs, temporal efficiency, and sociogeographic moderation; contribution to evidence synthesis, manuscript drafting, critical revision.

Email: nicotamarin2026@udec.cl