

Making Student Judgement Visible in AI-Assisted Feedback: A Decision Record for Higher Education

Dr. José Alberto Sotelo Martín¹, Dra. María Raquel Armas
Zavaleta²

¹,Universidad Internacional de La Rioja

²,Universidad Antonio de Nebrija

ABSTRACT *Generative artificial intelligence can provide extensive comments on university assignments, but a revised submission offers limited evidence of how students evaluated those comments. This conceptual article proposes a feedback decision record that links a selected recommendation to an assessment criterion, independent evidence, a student decision and a subsequent action. Four response options are distinguished: acceptance, adaptation, rejection and deferral. The record is situated within scholarship on feedback literacy, evaluative judgement and learning with artificial intelligence. Its specific contribution is an instructional procedure for discussing consequential revision decisions rather than inferring learning from the quantity of feedback or the polish of a final text. Constructed examples illustrate its use in psychology and education assignments. The article also identifies conditions under which the procedure could become burdensome or encourage the outsourcing of reflective explanations. A prospective evaluation should examine the quality of students' justifications and their performance on a subsequent task without AI assistance. The proposal offers a testable approach to preserving student responsibility for revision, while making no claim of demonstrated effectiveness.*

KEYWORDS *Artificial intelligence, Evaluative judgement, Feedback literacy, Higher education, Student agency*

I. Introduction

A student who receives detailed comments on an assignment still has to decide whether those comments are accurate, relevant and worth acting upon. University feedback therefore involves an educational relationship between information, judgement and action. Boud and Molloy [1] emphasise the design of feedback processes that develop learners' capabilities beyond the immediate task. Tai and colleagues [2] identify evaluative judgement as the capacity to make decisions about the quality of one's own and others' work. These perspectives direct attention to what the learner does with advice.

Generative artificial intelligence introduces a practical difficulty into this relationship. It can participate in producing comments, rewriting an assignment and explaining the changes. A teacher may consequently see an improved text and an articulate reflection without being able to determine whether the student understands either. Feedback literacy, as developed by Carless and Boud [3], provides a useful foundation for discussing students' capacity to interpret feedback and use it productively. In an AI-assisted environment, educators also need a manageable way to examine the decisions connecting comments with revision.

Thamdee and Man-khongdi [4] discuss conceptual gaps in AI-assisted feedback for English writing, including questions about deeper learning and learner autonomy. The present article focuses on disciplinary assignments in psychology and education, where a linguistic improvement may conceal a substantive error in interpretation. Its guiding question is how teachers can make consequential student judgements available for discussion when AI participates in feedback and revision.

This difficulty does not arise only when feedback is inaccurate. Sound advice can also be applied mechanically. A student may replace a claim with the correct qualification yet remain unable to explain why the qualification matters. Equally, the learner may understand a problem but struggle to express the response in polished academic language. For psychology and education programmes, these possibilities matter because written assignments often require the interpretation of evidence and the justification of professional choices. A teaching procedure should help distinguish the quality of the submitted product from the reasoning available to the student.

Winstone and colleagues [5] review interventions supporting learners' active engagement with feedback and conceptualise responsibilities within feedback dialogue. This work makes clear that receiving information and using it are different educational activities. The present proposal addresses the interval between them. It asks students to make a small number of important choices discussable, while recognising that a written explanation is only one source of information about understanding. The record is intended for ordinary formative teaching and can accompany comments from teachers or peers as well as AI.

The proposed response is a short feedback decision record. For selected comments, a student identifies the relevant criterion, checks the recommendation against an independent source or task requirement, decides how to respond, and explains the resulting action. The contribution is a specific instructional procedure situated within existing scholarship. It does not claim that reflective revision or student agency are new concepts, or that a record can establish authorship by itself. Its purpose is to support a conversation about reasons and provide a basis for future evaluation.

II. Conceptual approach and relationship to existing work

This article uses a conceptual design approach informed by targeted consultation of published scholarship and institutional guidance. Sources were located through public web searches on 3 October 2026, using combinations of generative AI, university feedback, student agency, feedback literacy and evaluative judgement. Foundational feedback publications were consulted alongside recent work to clarify the proposal's relationship to established concepts. Bibliographic details and accessible primary texts or author abstracts were checked against publisher, university repository and institutional records. This was a purposive source consultation, without a comprehensive database search, formal screening procedure or risk-of-bias assessment. It cannot support estimates of effectiveness or claims that no similar instructional procedure exists.

Three comparisons help specify the contribution. Zhan and colleagues [6] situate engagement with generative AI within a cycle of feedback forethought, control and retrospect. Lindsay and colleagues [7] examine the abundance of AI-generated feedback through its volume, availability, relevance and character. Lu and Nguyen [8] conceptualise learner autonomy in terms of responsibilities that persist when AI participates in academic work. The present proposal takes a narrower unit of analysis: the student's decision about an individual recommendation. It offers a way to document that decision for formative discussion within these broader perspectives.

The reasoning proceeds from an educational aim to a possible classroom procedure. If students are expected to exercise judgement, they need opportunities to relate recommendations to criteria and evidence. If teachers are to discuss that judgement, some account of the reasoning must be available. A concise record could supply such an account, provided it is combined with dialogue and subsequent learning tasks. This is a design argument, not evidence that documentation causes improved judgement. The proposed mechanism could fail if students treat the record as an administrative requirement or outsource its explanations.

No participants were recruited and no empirical dataset was analysed. The examples below are constructed instructional illustrations. The procedure and its assessment prompts are author-proposed tools, not validated instruments. Keeping these distinctions explicit allows a useful conceptual contribution without presenting a prospective activity as a completed study.

Recent reviews strengthen the rationale while also delimiting it. Lee and Moore [9] synthesised ten empirical studies of generative AI for automated feedback in higher education, identifying opportunities for integration with human instruction. Ba and colleagues [10] reviewed 129 articles on AI-assisted feedback published between 2014 and 2023 and stressed the need to combine technological and pedagogical perspectives. Their evidence extends beyond present-day conversational systems. Zhang, Li and Huang [11] mapped 51 empirical studies of generative AI feedback from 2023 to 2025. Taken together, these reviews show an expanding field with varied contexts and outcomes; they do not establish the effectiveness of the procedure proposed here.

A particularly relevant experiment by Farrokhnia and colleagues [12] compared teacher feedback with two AI prompting conditions among 70 graduate students. Higher rated AI feedback in one condition did not produce significantly greater revision improvements, and teacher feedback supported comparable gains. This finding

concerns a particular argumentative writing task, rather than a universal ranking of feedback sources. It nonetheless gives a direct empirical reason to examine the learner's uptake rather than assume that better comments necessarily produce better revision. The decision record is proposed as one possible scaffold for that uptake, not as an intervention tested in that experiment.

Deng and colleagues [13] synthesised experimental evidence on ChatGPT and student learning and recommended distinguishing output quality from demonstrated skills, examining longer-term outcomes and complementing subjective reports with objective measures. These recommendations inform the prospective evaluation below. They also constrain the article's argument: neither positive average findings about AI-supported learning nor successful individual systems justify assuming that any use of AI feedback will benefit students. The educational design, the comparison condition and the outcome being measured must remain explicit. The literature therefore supports a researchable question rather than a claim of established effectiveness.

III. Judgement and responsibility during revision

The proposed procedure distinguishes a feedback comment from a revision decision. A comment recommends or questions something. A decision identifies what the student will do and why. A changed sentence can be appropriate while the reasoning behind it remains weak. Conversely, a student may reasonably retain an original formulation when a recommendation conflicts with the evidence or the purpose of the task. Teachers therefore need to consider the justification and the work together.

Nicol and Macfarlane-Dick [14] relate formative assessment and feedback to students' regulation of their own learning. For the present design, this means giving the student a criterion and an opportunity to make a consequential choice. The student is not expected to accept every comment to demonstrate engagement. Nor is disagreement automatically a sign of independent thinking. An adequate response should identify the issue, examine relevant evidence and show how the decision affects the assignment.

Bearman and Ajjawi [15] discuss pedagogy for working with AI systems whose operations cannot be fully inspected. Their treatment of indeterminacy helps locate the teacher's task: students must learn to evaluate contributions from systems they cannot simply assume to be authoritative. In a decision record, a model's confident explanation is therefore insufficient evidence for a disputed disciplinary claim. The student should identify an appropriate reading, source passage, rubric requirement or other defensible basis for the choice.

The mechanism also draws on earlier accounts of formative assessment. Sadler [16] connects improvement with an understanding of quality, comparison of current performance against an appropriate standard, and the ability to act on the difference. Nicol [17] develops the role of internal feedback generated through comparison processes. A decision record translates these concerns into an explicit comparison between a recommendation and a disciplinary basis for judging it. The written record is not itself the internal feedback: it is a partial account that can help a teacher and student examine the comparison and identify where further understanding is needed.

Hattie and Timperley [18] distinguish feedback directed at the task, the process, self-regulation and the person. This distinction helps teachers select recommendations worth recording. Correcting an isolated typographical error may need little discussion, whereas reconsidering an inference or the organisation of evidence may expose a consequential learning issue. The same assignment can contain both. Selection should therefore depend on the learning objective and the decision's significance, rather than the length of the comment or the authority attributed to its source. Students should be taught this selection process instead of being expected to infer it from the form.

The distinction between assisted performance and subsequent learning also warrants attention. Bastani and colleagues [19] report a field experiment in high-school mathematics in which unrestricted AI assistance could improve practice performance while undermining later performance without assistance, with different outcomes under a safeguarded tutor design. This evidence concerns a different educational setting and task. It does not establish the effectiveness of decision records in university writing. It nevertheless supports treating immediate improvement and transferable learning as separate outcomes when evaluating the proposed procedure.

Evidence about AI should also accommodate successful instructional designs. Kestin and colleagues [20] found greater short-term learning with a purpose-built AI tutor than with an active-learning class in a university physics experiment. This result and the mathematics findings above concern different systems, learners and comparisons, so their effects cannot be directly contrasted. Together they suggest that pedagogical constraints and the conditions of assessment matter. For the present proposal, the relevant question is whether selected justification and dialogue add value to a specified feedback activity, not whether AI is beneficial or harmful in the abstract.

IV. The feedback decision record

A record concerns one recommendation that could materially change an assignment's argument, interpretation, evidence or organisation. Students need not reproduce every interaction with an AI system. A teacher can initially request two consequential decisions from a revision task and explain why those choices matter. This suggested number is a practical starting point, not an empirically established optimum. An activity with one substantial comment may require only one record.

Each record contains six fields. The first identifies the recommendation, through a short quotation or an accurate summary. The second names the criterion or purpose to which it relates. The third identifies the evidence used to evaluate it, with enough detail for the teacher to locate the material. The fourth states the decision. The fifth describes the action taken, including the location of any revision. The sixth gives the student's reason and identifies remaining uncertainty. A single concise paragraph can supply these fields when a form would interrupt the task.

The four available decisions are acceptance, adaptation, rejection and deferral. Acceptance means that the recommendation is followed for a stated reason. Adaptation means that its underlying concern is addressed through a different change. Rejection means that the recommendation is not adopted because it conflicts with the evidence, criterion or defensible purpose of the work. Deferral means that the student cannot yet resolve the issue and identifies a check or conversation needed before acting. These options describe responses, not levels of competence. The quality of a decision depends on its justification.

Deferral is especially useful when a student lacks the basis to assess an apparently authoritative comment. Requiring an immediate acceptance or rejection could encourage a confident explanation where uncertainty would be more appropriate. A deferred decision should name the unresolved question and the next step. It should not become an indefinite excuse to avoid revision. The teacher may resolve a task requirement directly or identify a reading that allows the student to reconsider the recommendation.

Evidence independent of the disputed recommendation need not mean an external publication for every decision. A methodological claim may require the original study; a question about the task's scope may be resolved through its instructions; an organisational choice may be justified through the argument's structure and an agreed criterion. The check must be suitable for the issue. Asking the same model to confirm its own advice would not provide an independent disciplinary basis. An additional AI response can suggest a check, but the student should be able to locate and discuss the material on which the final judgement rests.

The distinction between evidence and reason prevents the form from becoming a list of citations. Evidence identifies the material consulted; the reason explains how that material bears on the recommendation. A rubric saying that an argument should be clear does not, by itself, justify any proposed rewrite. The student must explain what was unclear and why the chosen action improves it. Where several interpretations are defensible, the record can identify the trade-off, such as preserving methodological caution while improving readability. This permits a proportionate account of judgement without implying that every revision has one correct answer.

Table 1 Response options within the feedback decision record

Decision	Reason to choose it	Action to document
Acceptance	The advice fits the criterion and evidence.	Identify the revision and why it is warranted.
Adaptation	The concern is valid but the proposed solution needs changing.	Explain the alternative revision.
Rejection	The advice conflicts with evidence or the task.	Explain why the original choice is retained.
Deferral	The available basis is insufficient to decide.	Name the unresolved issue and the next check.

V. Constructed examples in psychology and education

Consider an undergraduate psychology assignment discussing an association between social isolation and wellbeing. An AI comment recommends replacing cautious wording with a direct causal claim to make the conclusion stronger. The student checks the design described in the assigned study and finds that the evidence is correlational. A justified rejection would preserve the qualification and explain why the proposed wording would exceed the evidence. The instructional value lies in the student's ability to relate language to research design. The example does not report the output of an actual model or the behaviour of an actual student.

In a second illustration, an education assignment describes a conflict between pupils and a proposed classroom response. A comment asks for greater specificity. The student accepts the concern but adapts the suggestion by describing how the activity relates to the course's stated objective, rather than adding unsupported claims about outcomes. The record points to the rubric requirement and the paragraph revised. The teacher can then discuss whether the new detail addresses the objective or merely lengthens the text.

A third illustration concerns a recommendation to add a reference. The student cannot locate the proposed publication in a publisher or library record. Deferral identifies the uncertainty and the planned check. The student can subsequently reject the reference if it cannot be verified and use an appropriate source already available. The pedagogical question is whether the student can explain what the source should support. Successfully locating a citation would not by itself establish that it supports the argument.

These examples allow educators to model distinct responses before students apply the procedure to their own assignments. All students can receive the same teacher-constructed comments, including an ambiguous or misleading suggestion clearly identified as instructional material. This provides an opportunity to practise judgement without requiring paid access to a particular AI product. Examples should remain open to defensible disagreement where a task permits more than one interpretation.

A completed record for the psychology illustration might state: the recommendation is to describe isolation as causing poorer wellbeing; the criterion is accurate interpretation of research design; the evidence is the assigned study's methods and its stated limitations; the decision is rejection; the action is to retain associative wording and clarify the design; the reason is that the study does not establish the proposed causal relationship. This is a constructed example, not a quotation from a participant. During discussion, the teacher could ask what kind of evidence would warrant stronger language, moving from a local revision towards understanding the criterion.

Walters and Wilder [21] documented fabricated references and bibliographic errors in outputs from the model versions they studied. Their findings do not provide an error rate for current systems, but they support treating bibliographic verification as a separate task. In the citation illustration, the learner should check both existence and relevance: a genuine article may still fail to support the assertion for which it is proposed. A useful record names the passage or finding that supports the claim, rather than reporting only that a DOI resolves. This makes source evaluation available for discussion alongside citation accuracy.

In the education illustration, adaptation becomes defensible when the student identifies a specific mismatch between the suggested activity and the assignment's objective. The record can state which element was retained, which was changed and what criterion guided that choice. A teacher might ask whether the alternative would remain appropriate for a different pupil group. This question exposes the limits of the original justification without pretending that a written plan demonstrates effectiveness in a real classroom. Across the examples, the educational demand remains consistent: the student must connect a decision to reasons that can be examined and reconsidered.

VI. Classroom use and formative assessment

Before revision, the teacher explains the assignment criteria and demonstrates a decision record using a short example. Students then identify selected recommendations and prepare their own records alongside the revised text. After submission, the teacher chooses one decision for a brief written exchange, seminar discussion or tutorial question. A question such as "What evidence would change your decision?" can reveal whether the explanation rests on a reason the student can use beyond the original wording.

Feedback on the record should examine the relevance of the criterion, the adequacy of the evidence and the coherence between the decision and the revision. An unsupported explanation might repeat the recommendation without explaining its merits. A more developed explanation identifies relevant material but does not resolve the specific issue. A well-supported explanation applies that material to the issue and recognises a relevant limitation. These descriptions are proposed formative prompts. They are not a validated scoring scale and should not be turned into numerical grades without further development.

The teacher should avoid rewarding a particular pattern of decisions. Requiring one acceptance, one adaptation and one rejection would distort the activity if the available advice did not warrant those responses. Similarly, a high proportion of rejected comments would not prove student independence. A student who accepts sound advice after an appropriate check may exercise better judgement than one who rejects it to appear critical. Discussion should examine reasons and consequences.

A subsequent task is needed to explore transfer. After revising an argument with assistance, students could evaluate a short new argument against the same disciplinary criterion without AI. Their explanation should be judged independently of the polish of the revised assignment. This step is a proposed evaluation opportunity rather than proof of retention. It also helps teachers identify cases in which students can reproduce a record but struggle to apply the underlying distinction elsewhere.

Peer exchange can provide an additional comparison before teacher discussion. Banihashem and colleagues [22] compared AI and peer feedback on argumentative essays and found different characteristics that suggested a complementary role. This study evaluated feedback characteristics, not the proposed decision record. In classroom use, pairs could examine one another's evidence and ask whether the reason actually supports the action. The teacher would remain responsible for clarifying disputed disciplinary requirements. Peer discussion should not require students to disclose complete AI conversations or personal information; sharing the selected decision and relevant source passage is sufficient for the proposed exercise.

A manageable sequence would reserve a short part of a revision session for selecting and checking comments, followed by a focused discussion in an existing seminar or tutorial. The time allocation should be adjusted locally rather than imposed as a fixed dose. At first, the teacher can supply a criterion and a worked example; later, students can justify which recommendation deserves attention. This gradual reduction of support gives the activity a developmental purpose. The same six fields can be retained across tasks while the complexity of the evidence and the learner's responsibility for selection increase.

VII. Constraints and conditions for responsible use

Documentation can displace learning if it becomes too elaborate. Requiring complete conversational histories, extensive logs or separate explanations for minor stylistic changes could impose considerable work on both students and teachers. The proposal therefore concentrates on selected decisions with substantive consequences. A formative trial should examine the time required and whether the information produced actually improves feedback conversations.

The record is also vulnerable to automated completion. A student can ask a model to justify a choice after the revision has already been made. The resulting text may resemble a thoughtful explanation without reflecting the student's understanding. No written form can eliminate that possibility. A brief follow-up conversation, comparison with the relevant source and a later task can provide additional evidence, but they cannot establish complete certainty about the origins of every sentence. The record should support learning rather than operate as an authorship detector.

Accessibility requires more than allowing everyone to use the same form. Some students may communicate their reasoning more clearly through speech, annotated passages or a short supported conversation. Alternative formats should preserve the same educational demand while avoiding unnecessary language or transcription barriers. An equivalent route using teacher or peer comments should be available where students lack suitable AI access or prefer not to use it.

Miao and Holmes [23] frame educational use of generative AI around human agency, institutional capacity and responsible design, including attention to privacy. For a decision-record activity, teachers can apply this orientation by using approved systems where appropriate and avoiding the upload of identifying or confidential material. In professional programmes, anonymisation of a text should not be assumed to make a real client record suitable for an external platform. Fictional educational examples offer a simpler basis for practising the procedure.

Risko and Gilbert [24] describe cognitive offloading as using external action to reduce the demands placed on internal processing. Their review predates contemporary generative AI and should not be read as evidence that using a model necessarily weakens thinking. External support may release capacity for demanding reasoning. The concern here is the specific delegation of the judgement that the assignment is meant to develop. If the system supplies the recommendation, revision and justification, students may have fewer opportunities to practise that judgement. The record therefore needs to be taught as an activity of checking and deciding, not merely producing another text.

Kasneci and colleagues [25] discuss educational opportunities alongside challenges of large language models, while Yan and colleagues [26] identify concerns about transparency, replicability and privacy in educational applications. These analyses support specifying local conditions of use. Teachers should make permitted assistance understandable, provide an equivalent route with human comments, and explain what information may be shared with an approved system. Recording the source and relevant date of an AI comment can assist contextual interpretation, but the teaching form should not become an exhaustive technical log. Research evaluation requires more detailed documentation of the model and instructions.

VIII. Propositions for future evaluation

Two propositions make the proposal open to examination. First, requiring an evidence-linked decision could improve the quality of students' explanations compared with revising from comments alone. Second, discussing a selected decision could improve performance on a subsequent task that asks students to apply the same criterion without AI. Both propositions concern plausible educational mechanisms. Neither is established by the conceptual analysis presented here.

A prospective study could compare an AI-feedback activity with an otherwise similar activity that adds the decision record. A further comparison could add the follow-up discussion, allowing the contribution of documentation and dialogue to be examined separately. Teachers would need to specify the task, criteria, permitted assistance and time available. A common set of comments could help control variation in the advice received during an initial evaluation, although later work should examine authentic student-requested feedback.

The primary outcome should concern judgement rather than satisfaction with the tool. Possible outcomes include the quality of the justification and performance on a new unassisted task. Independent raters could assess anonymised responses against a developed rubric, with an agreement analysis appropriate to its scale. Secondary outcomes could include time burden, perceived usefulness and accessibility. Prior attainment, disciplinary knowledge and familiarity with AI should be considered when planning allocation and analysis. Sample size and analysis decisions would require a justified protocol before recruitment.

A successful evaluation would also need to look for adverse effects. The record might increase anxiety, encourage performative explanations or make revision less efficient. Qualitative accounts could identify when students found the activity useful and when it became compliance work. Findings should distinguish a locally workable teaching routine from a generalisable improvement in evaluative judgement. Any research involving students would require the institutional review and consent arrangements applicable to that setting.

An initial development phase should establish whether students and raters interpret the proposed fields consistently. Think-aloud activities with a small, appropriately consented group could reveal whether the evidence field prompts checking or merely citation copying. Draft scoring descriptors should distinguish accurate application of a criterion from fluent but irrelevant explanation. Raters can discuss examples before independent assessment, and disagreements can inform rubric revision. This preparatory work would support the interpretation of later scores; it would not, by itself, validate a general measure of student agency or prove that the activity improves learning.

The transfer task should preserve the criterion while changing the example enough to prevent simple reproduction of the original explanation. A delayed assessment could examine whether the distinction remains available after the revision session. Research reports should describe assistance restrictions, missing responses, allocation procedures and any clustering within teaching groups. They should also report uncertainty around estimates and avoid treating a non-significant difference as proof of equivalence. These decisions require a prospective protocol and feasible staffing. The aim is to assess a defined teaching procedure under stated conditions, with conclusions proportionate to the evidence.

IX. Conclusion

The feedback decision record proposes a concrete way to discuss student judgement when AI participates in university revision. Its six fields connect a recommendation with a criterion, evidence, decision, action and reason. Acceptance, adaptation, rejection and deferral allow students to respond to the quality of the advice rather than demonstrate engagement through automatic compliance. The procedure is especially relevant to assignments in which changes to wording can alter the interpretation of evidence.

The proposal's value depends on how it is taught and evaluated. A concise record combined with discussion could make important decisions available for formative attention. A lengthy administrative exercise or an AI-generated justification could obscure them. Future work should therefore examine the procedure's burden and its

contribution to students' performance on subsequent tasks. Until such evidence is available, it should be treated as a testable instructional design rather than an established intervention.

X. Use of generative artificial intelligence

Generative artificial intelligence assisted with initial drafting, language editing, preparation of the Spanish version and discovery of sources for this conceptual manuscript. No participant records, empirical observations or statistical results were generated. The instructional examples are constructed illustrations and are identified as such.

References

- [1] D. Boud and E. Molloy, Rethinking models of feedback for learning: the challenge of design, *Assessment & Evaluation in Higher Education*, 38(6), 2013, 698-712. <https://doi.org/10.1080/02602938.2012.691462>
- [2] J. Tai, R. Ajjawi, D. Boud, P. Dawson and E. Panadero, Developing evaluative judgement: enabling students to make decisions about the quality of work, *Higher Education*, 76, 2018, 467-481. <https://doi.org/10.1007/s10734-017-0220-3>
- [3] D. Carless and D. Boud, The development of student feedback literacy: enabling uptake of feedback, *Assessment & Evaluation in Higher Education*, 43(8), 2018, 1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
- [4] N. Thamdee and P. Man-khongdi, Beyond Surface Integration: Mapping the Conceptual Gaps and Theoretical Frameworks of AI-Assisted Feedback in English Writing, *International Journal of Arts and Social Science*, 9(2), 2026, 244-260. <https://www.ijassjournal.com/2026/V9I2/51466642241.pdf>
- [5] N. E. Winstone, R. A. Nash, M. Parker and J. Rowntree, Supporting learners' agentic engagement with feedback: a systematic review and a taxonomy of recipience processes, *Educational Psychologist*, 52(1), 2017, 17-37. <https://doi.org/10.1080/00461520.2016.1207538>
- [6] Y. Zhan, D. Boud, P. Dawson and Z. Yan, Generative artificial intelligence as an enabler of student feedback engagement: a framework, *Higher Education Research & Development*, 44(5), 2025, 1289-1304. <https://doi.org/10.1080/07294360.2025.2476513>
- [7] E. Lindsay, A. Rodda, A. Lidfors Lindqvist, Z. Quince, M. Lim and D. Jiang, The Dimensions of Abundance in AI-Generated Feedback, *Education Sciences*, 16(3), 2026, 465. <https://doi.org/10.3390/educsci16030465>
- [8] Q. V. Lu and N. C. L. Nguyen, Reconceptualising Learner Autonomy in Generative AI-Supported Higher Education: A Technology-Mediated Framework, *International Journal of Learning, Teaching and Educational Research*, 25(9), 2026, 523-550. <https://doi.org/10.26803/ijlter.25.9.21>
- [9] S. S. Lee and R. L. Moore, Harnessing Generative AI (GenAI) for Automated Feedback in Higher Education: A Systematic Review, *Online Learning*, 28(3), 2024, 82-104. <https://doi.org/10.24059/olj.v28i3.4593>
- [10] S. Ba, L. Yang, Z. Yan, C. K. Looi and D. Gašević, Unraveling the mechanisms and effectiveness of AI-assisted feedback in education: A systematic literature review, *Computers and Education Open*, 9, 2025, 100284. <https://doi.org/10.1016/j.caeo.2025.100284>
- [11] K. Zhang, S. Li and Y. Huang, Generative AI-driven feedback in higher education: a scoping review, *Assessment & Evaluation in Higher Education*, 2026, advance online publication. <https://doi.org/10.1080/02602938.2026.2617157>
- [12] M. Farrokhnia, S. Latifi, P. M. Papadopoulos, L. Hogenkamp, H. Gijlers, H. Khosravi and O. Noroozi, Generative AI offers more, but students revise less: comparing the effects of teacher and AI feedback on student essay revisions, *International Journal of Educational Technology in Higher Education*, 23, 2026, 6. <https://doi.org/10.1186/s41239-026-00579-9>
- [13] R. Deng, M. Jiang, X. Yu, Y. Lu and S. Liu, Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies, *Computers & Education*, 227, 2025, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- [14] D. J. Nicol and D. Macfarlane-Dick, Formative assessment and self-regulated learning: a model and seven principles of good feedback practice, *Studies in Higher Education*, 31(2), 2006, 199-218. <https://doi.org/10.1080/03075070600572090>

-
- [15] M. Bearman and R. Ajjawi, Learning to work with the black box: Pedagogy for a world with artificial intelligence, *British Journal of Educational Technology*, 54(5), 2023, 1160-1173. <https://doi.org/10.1111/bjet.13337>
- [16] D. R. Sadler, Formative assessment and the design of instructional systems, *Instructional Science*, 18, 1989, 119–144. <https://doi.org/10.1007/BF00117714>
- [17] D. Nicol, The power of internal feedback: exploiting natural comparison processes, *Assessment & Evaluation in Higher Education*, 46(5), 2021, 756–778. <https://doi.org/10.1080/02602938.2020.1823314>
- [18] J. Hattie and H. Timperley, The power of feedback, *Review of Educational Research*, 77(1), 2007, 81–112. <https://doi.org/10.3102/003465430298487>
- [19] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakçı and R. Mariman, Generative AI without guardrails can harm learning: Evidence from high school mathematics, *Proceedings of the National Academy of Sciences*, 122(26), 2025, e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- [20] G. Kestin, K. Miller, A. Klales, T. Milbourne and G. Ponti, AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting, *Scientific Reports*, 15, 2025, 17458. <https://doi.org/10.1038/s41598-025-97652-6>
- [21] W. H. Walters and E. I. Wilder, Fabrication and errors in the bibliographic citations generated by ChatGPT, *Scientific Reports*, 13, 2023, 14045. <https://doi.org/10.1038/s41598-023-41032-5>
- [22] S. K. Banihashem, N. T. Kerman, O. Noroozi, J. Moon and H. Drachsler, Feedback sources in essay writing: peer-generated or AI-generated feedback?, *International Journal of Educational Technology in Higher Education*, 21, 2024, 23. <https://doi.org/10.1186/s41239-024-00455-4>
- [23] F. Miao and W. Holmes, *Guidance for generative AI in education and research* (Paris: UNESCO, 2023). <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- [24] E. F. Risko and S. J. Gilbert, Cognitive offloading, *Trends in Cognitive Sciences*, 20(9), 2016, 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>
- [25] E. Kasneci et al., ChatGPT for good? On opportunities and challenges of large language models for education, *Learning and Individual Differences*, 103, 2023, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- [26] L. Yan, L. Sha, L. Zhao, Y. Li, R. Martinez-Maldonado, G. Chen, X. Li, Y. Jin and D. Gašević, Practical and ethical challenges of large language models in education: A systematic scoping review, *British Journal of Educational Technology*, 55(1), 2024, 90–112. <https://doi.org/10.1111/bjet.13370>