International Journal of Essential Competencies in Education
https://journal-center.litpam.com/index.php/ijece/index

June 2026 Vol. 5, No. 1
e-ISSN: 2962-2131
pp. 1-12

When Better Feedback Is Not Enough: Why GenAI-Supported Peer Assessment Needs Self-Reflection to Foster Preservice Teachers' Critical Thinking

Ni Nyoman Sri Putu Verawati*, Wahyudi, Nina Nisrina

Physics Education Department, University of Mataram, Mataram, Indonesia

*Correspondence: veyra@unram.ac.id

Article Info	Abstract

Article History
Received: March 2026;
Revised: April 2026;
Published: June 2026

Keywords
Generative artificial intelligence;
Peer assessment;
Self-reflection;
Feedback uptake;
Critical thinking

Generative artificial intelligence is increasingly used to support feedback, assessment, and learning in teacher education. This commentary examines a recent empirical study on GenAI-supported peer feedback for preservice teachers and argues that its most important contribution lies not only in showing that AI can improve feedback quality, but also in demonstrating the central role of self-reflection in transforming better feedback into deeper thinking. The focal study reports positive effects on feedback quality, technology-enhanced learning design artifacts, feedback uptake, and critical thinking tendency. However, the evidence also suggests that improved feedback does not automatically produce critical thinking. The effect becomes more educationally meaningful when learners apply self-feedback, internalize evaluative criteria, and use peer assessment as an opportunity for reflective revision. This commentary discusses the strengths of the focal study, offers a cautious interpretation of its findings, and identifies several limitations related to short intervention duration, quasi-experimental design, self-reported critical thinking measures, and the unresolved role of GenAI in verifying feedback accuracy. It concludes that GenAI-supported peer assessment should be designed not as a shortcut to better comments, but as a scaffold for feedback literacy, self-regulation, self-reflection, and professional judgment in teacher education.

https://doi.org/10.36312/xx4deg57

How to Cite

Verawati, N. N. S. P., Wahyudi, W., & Nisrina, N. (2026). When Better Feedback Is Not Enough: Why GenAI-Supported Peer Assessment Needs Self-Reflection to Foster Preservice Teachers' Critical Thinking. *International Journal of Essential Competencies in Education*, *5*(1), 1-12. https://doi.org/10.36312/xx4deg57

INTRODUCTION

Critical thinking remains one of the central aims of higher education and teacher education. It is commonly understood as purposeful, reflective, and evidence-based judgment that involves interpretation, analysis, evaluation, inference, explanation, and self-regulation (Facione, 1990; Ennis, 2018; Facione, 2020). This means that critical thinking is not simply a matter of "thinking deeply" or expressing opinions. It requires learners to examine evidence, test assumptions, judge the quality of arguments, consider alternatives, and revise conclusions when better reasons emerge (Halpern, 2014; Niu et al., 2013). For preservice teachers, this competence is especially important because teaching requires continuous judgment about content, pedagogy, learners' needs, classroom evidence, and instructional decisions (Barahona et al., 2023; Liu et al., 2018; Lu et al., 2024).

Research on critical thinking has repeatedly shown that this competence does not develop automatically through ordinary instruction. Meta-analytic evidence indicates that critical thinking is more likely to improve when instruction explicitly targets critical thinking skills and dispositions rather than assuming that they will emerge as a by-product of general learning activities (Abrami et al., 2008; Abrami et al., 2015; Niu et al., 2013). This point is important for teacher education because preservice teachers may complete complex tasks, use digital tools, or participate in peer assessment without necessarily developing stronger critical reasoning. The learning environment must be intentionally designed to require analysis, evaluation, justification, revision, and reflection.

The rapid development of generative artificial intelligence has intensified this issue. In AI-mediated learning environments, students and future teachers increasingly encounter machine-generated suggestions, automated feedback, and algorithmic judgments. These resources can support learning, but they also require careful evaluation because GenAI may produce plausible yet inaccurate, incomplete, or context-insensitive output (Bozkurt et al., 2024; Feldman-Maggor et al., 2025; Jukiewicz, 2024; Usher, 2025). Therefore, teacher education needs to prepare preservice teachers not only to use GenAI efficiently, but also to evaluate it critically and pedagogically.

Peer assessment has long been regarded as a promising strategy for fostering critical thinking. When learners assess peers' work, they are required to compare performance with criteria, identify strengths and weaknesses, justify judgments, and provide suggestions for improvement (Topping, 2009; van Popta et al., 2017; Zhan et al., 2023). In online peer assessment, learners take dual roles as assessors and assessees, which creates opportunities for learning both from receiving feedback and from giving feedback (Cho & Cho, 2011; Fang et al., 2021; Li et al., 2010; Zhang et al., 2021). However, the potential of peer assessment depends strongly on whether students can produce meaningful feedback and use it in subsequent revisions.

Classic work on formative assessment emphasizes that feedback becomes useful when learners understand the standard of quality, compare their current performance with that standard, and act to close the gap (Sadler, 1989). Similarly, effective feedback should help learners understand where they are going, how they are progressing, and what they need to do next (Hattie & Timperley, 2007). In critical thinking instruction, this means that feedback should not merely tell learners whether their work is correct or incorrect. It should help them understand why an argument is weak, what evidence is missing, how reasoning can be strengthened, and what revision would improve the quality of judgment.

The quality of feedback matters because vague comments such as "be more critical" or "improve your argument" rarely support meaningful revision. Formative feedback is more useful when it is specific, timely, supportive, and focused on the process of improvement (Shute, 2008). In peer assessment, high-quality feedback should provide actionable information that helps the recipient revise the work and helps the feedback provider sharpen evaluative judgment (Banihashem et al., 2022; Carless & Boud, 2018; Valero Haro et al., 2024). Yet students often produce superficial praise, unsupported criticism, or general suggestions that are difficult to use (Banihashem et al., 2024; Gao et al., 2024; Kerman et al., 2024). This is one reason why GenAI-supported peer feedback has become an attractive instructional possibility.

GenAI can support peer feedback by reviewing students' feedback drafts, identifying weaknesses, and suggesting ways to make comments more specific, constructive, and actionable (Guo, 2024; Guo et al., 2024; Guo et al., 2025; Noroozi et al., 2025b). However, the educational value of GenAI-supported feedback should not be reduced to producing better text. Better feedback is only useful when learners can interpret it, evaluate it, and act on it. This is the core concern of feedback literacy, which refers to students' capacity to understand, make sense of, and use feedback productively (Carless & Boud, 2018; Han & Xu, 2020; Winstone et al., 2017). It is also connected to self-regulated learning, because learners must monitor their performance, evaluate strategies, and adjust their work based on feedback (Butler & Winne, 1995; Nicol & Macfarlane-Dick, 2006).

This commentary discusses the focal empirical study by Zhou et al. (2026), which examined the effects of GenAI-supported peer feedback on preservice teachers' learning outcomes, feedback uptake, and critical thinking. The article offers a useful model for integrating GenAI into online peer assessment in teacher education. It shows that GenAI-supported peer feedback can improve cognitive feedback quality, support technology-enhanced learning design artifact revision, increase feedback uptake, and foster critical thinking tendency (Zhou et al., 2026). At the same time, this commentary argues that the study's strongest implication is not that GenAI directly produces critical thinking. Rather, GenAI may support critical thinking when it helps learners internalize evaluative criteria, generate self-feedback, and engage in reflective revision.

COMMENTARY ON THE FOCAL ARTICLE

The focal article addresses a timely problem in teacher education: how to improve peer feedback so that peer assessment becomes more useful for learning and critical thinking. Peer feedback is often promoted because it increases opportunities for evaluation, dialogue, and reflection (van Popta et al., 2017; Zhan et al., 2023). However, the effectiveness of peer feedback depends on its quality and on whether learners take it up in subsequent revisions (Er et al., 2021; Wu & Schunn, 2020a; Wu & Schunn, 2020b). The focal study is therefore well positioned because it does not assume that peer feedback is automatically effective. Instead, it asks how GenAI can help preservice teachers produce higher-quality feedback during online peer assessment.

The GenAI-supported peer feedback approach follows a human-AI collaboration model. Preservice teachers first submit technology-enhanced learning design artifacts through an online peer assessment platform. They then assess peers' artifacts using a rubric and draft feedback. In the experimental condition, the draft feedback is submitted to a GenAI tool, which critiques the comments and provides suggestions for improvement. Students revise their feedback before sending it to peers, and the recipients then use the improved feedback to revise their artifacts (Zhou et al., 2026). This design is pedagogically meaningful because GenAI is not positioned as a replacement for the human assessor. Instead, it serves as a scaffold for feedback improvement.

The study involved 199 preservice chemistry teachers enrolled in a Modern Educational Technology course. The experimental group used GenAI-supported peer feedback, while the control group used traditional peer feedback. Both groups completed peer assessment activities, revised TEL design artifacts, and completed pre- and post-questionnaires on critical thinking tendency (Zhou et al., 2026). The study measured feedback quality, improvement of TEL design artifacts, feedback uptake, and critical thinking tendency. This combination of variables is useful because it connects feedback production, feedback use, artifact revision, and thinking disposition.

The findings show that GenAI-supported peer feedback increased guidance-oriented cognitive feedback, improved TEL design artifacts across TPACK-related dimensions, increased the application of peer and self-feedback, and produced a statistically significant improvement in critical thinking tendency (Zhou et al., 2026). The moderation result is especially important. Applying self-feedback strengthened the relationship between GenAI-supported peer feedback and critical thinking tendency. This finding supports the argument that critical thinking may develop not merely because students receive better comments, but because they learn to apply evaluative criteria to their own work.

This point connects the focal study to broader research on evaluative judgement. Evaluative judgement refers to learners' capacity to make informed decisions about the quality of their own work and the work of others (Tai et al., 2018). This capacity is closely related to critical thinking because both require criteria-based judgment, evidence evaluation, comparison, and justification. Peer assessment can help develop evaluative judgement when students actively use rubrics, examine examples, compare quality, and revise their own work (Panadero & Jonsson, 2013; Tai et al., 2018). In this sense, the focal study's contribution is stronger when read through the lens of evaluative judgement and self-feedback, not simply through the lens of AI-assisted efficiency.

The study also has implications for internal feedback. Internal feedback occurs when learners compare their work with external information such as criteria, exemplars, peer comments, teacher comments, or previous performance, and then generate insights for improvement (Nicol, 2021). The focal study's finding on self-feedback uptake suggests that GenAI-supported peer assessment may support this comparison process. When students evaluate peers' work, improve their own comments with GenAI, and then apply similar criteria to their own artifacts, they are not merely using external feedback. They are generating internal feedback through comparison and reflection.

WHAT THE STUDY DOES WELL

The first strength of the focal study is its clear pedagogical design. GenAI is embedded within a structured peer assessment workflow involving rubrics, feedback drafting, AI-supported feedback revision, peer feedback delivery, and artifact revision (Zhou et al., 2026). This design matters because effective feedback environments require explicit standards, opportunities for comparison, and chances to act on feedback (Sadler, 1989; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006). Without such structure, GenAI use may become a shortcut for producing polished comments rather than a tool for developing evaluative reasoning.

The second strength is the focus on feedback uptake. Feedback does not improve learning simply because it is provided. Learners must understand it, judge its relevance, and use it in revision (Carless & Boud, 2018; Winstone et al., 2017; Wood, 2022). By coding uptake as ignore, copy, and apply, the focal study provides a more precise account of how students use feedback (Zhang et al., 2021; Zhou et al., 2026). This distinction is important because copying feedback is not the same as applying it thoughtfully. Critical thinking is more likely to be supported when learners adapt feedback through judgment rather than reproduce it mechanically.

The third strength is the distinction between peer feedback uptake and self-feedback uptake. Students can learn from feedback they receive, but they can also learn from feedback they provide to others (Cho & Cho, 2011; Li et al., 2010; van Popta et al., 2017). The focal study shows that applying self-feedback is central to the relationship between GenAI-supported peer feedback and critical thinking tendency (Zhou et al., 2026). This is theoretically important because it suggests that the process of judging others' work becomes more powerful when students turn the same criteria back on their own work.

The fourth strength is the use of TEL design artifacts as learning outcomes. The study evaluates improvements in artifacts using TPACK-related dimensions, which are relevant for preservice teachers' professional preparation (Schmid et al., 2021; Yeh et al., 2021). This provides stronger evidence than relying only on self-reported perceptions. Since teacher education should help students design meaningful learning with technology, artifact-based assessment offers a relevant way to evaluate learning outcomes (Chen et al., 2021; Zhang et al., 2024).

The fifth strength is the study's practical relevance for large-class teaching. GenAI-supported peer feedback may help address the difficulty of monitoring student feedback at scale (Guo, 2024; Guo et al., 2025). Instructors often cannot provide detailed feedback on every piece of peer feedback, especially in large online or blended classes. A structured GenAI-supported approach may therefore help students improve feedback quality while allowing instructors to focus on designing criteria, monitoring learning processes, and supporting reflective practice.

WHAT THE RESULTS SUGGEST

The results suggest that GenAI-supported peer feedback is most effective in improving cognitive feedback quality. The experimental group produced more guidance-oriented cognitive feedback than the control group (Zhou et al., 2026). This finding aligns with previous studies showing that AI-supported systems can help learners generate more specific and constructive feedback (Guo et al., 2024; Guo et al., 2025; Noroozi et al., 2025b). It also aligns with formative feedback theory, which emphasizes that feedback should give learners information about how to improve rather than merely evaluate performance (Sadler, 1989; Shute, 2008).

The findings also suggest that better feedback can support better artifact revision. Students in the GenAI-supported condition improved more across TPACK-related dimensions (Zhou et al., 2026). This supports the claim that actionable feedback can help learners close the gap between current performance and desired standards (Hattie & Timperley, 2007; Sadler, 1989). However, this finding should be interpreted as evidence of improved performance on TEL design artifacts, not as direct proof of broader teaching competence. A revised artifact may show learning within a specific task, but it does not necessarily demonstrate transfer to authentic classroom teaching.

The feedback uptake results are particularly important. Students in the GenAI-supported group were more likely to apply feedback and less likely to ignore it (Zhou et al., 2026). This supports the argument that feedback quality and feedback uptake are closely related (Wu & Schunn, 2020a; Wu & Schunn, 2020b). More specific and actionable feedback is easier for learners to interpret and use. Yet the more interesting finding is that self-feedback application plays a moderating role. This suggests that the educational value of GenAI-supported peer assessment depends on whether students use feedback processes to regulate their own learning.

The results on critical thinking tendency are promising but require caution. The focal study found a statistically significant improvement in critical thinking tendency in the GenAI-supported condition (Zhou et al., 2026). However, the evidence is based on self-report, and the effect is more modest than the effects on feedback behavior and artifact improvement. This distinction matters because critical thinking involves actual reasoning performance, not only perceived tendency (Facione, 1990; Ennis, 2018; Halpern, 2014). A stronger conclusion would be that GenAI-supported peer feedback can create conditions that support critical thinking, especially when students apply self-feedback, rather than claiming that GenAI directly produces critical thinking.

The moderation finding provides the strongest conceptual insight. Applying self-feedback strengthened the positive relationship between GenAI-supported peer feedback and critical thinking tendency (Zhou et al., 2026). This suggests that critical thinking is fostered through a reflective loop: learners evaluate others' work, refine their feedback with GenAI, compare evaluative criteria with their own artifact, and revise their own work. This process is consistent with self-regulated learning theory, which views feedback as information used for monitoring and adjustment (Butler & Winne, 1995; Nicol & Macfarlane-Dick, 2006). It is also consistent with internal feedback theory, where learners generate feedback for themselves through comparison (Nicol, 2021).

CRITICAL APPRAISAL AND ALTERNATIVE INTERPRETATIONS

The focal study makes an important contribution, but several interpretive risks should be considered. The first issue concerns the measurement of critical thinking. The study measured critical thinking tendency through a self-report questionnaire (Zhou et al., 2026). Such instruments can capture perceived disposition or awareness, but they do not directly measure critical thinking performance. Students may report stronger critical thinking because they feel more engaged, more confident, or more aware of criteria after using GenAI. This is meaningful, but it is not equivalent to demonstrating improved analysis, evaluation, inference, or self-regulation in authentic performance tasks (Facione, 1990; Ennis, 2018; Noroozi et al., 2025a).

The second issue concerns the short intervention duration. The study was conducted across a limited four-week process (Zhou et al., 2026). Short interventions can show positive effects, but they may also be influenced by novelty or temporary motivation (Lu & Ba, 2025). Since GenAI is still relatively new in many educational contexts, students may respond positively because the activity feels innovative or because AI support reduces the difficulty of writing feedback. Longer studies are needed to determine whether GenAI-supported peer assessment produces durable changes in feedback literacy, self-regulation, and critical thinking.

The third issue concerns the active ingredient of the intervention. The study attributes positive outcomes to GenAI-supported peer feedback, but the intervention includes several components: rubric-based evaluation, AI critique, feedback revision, peer feedback uptake, self-feedback uptake, and artifact revision (Zhou et al., 2026). It is therefore difficult to determine whether the improvement was caused by GenAI itself, by repeated use of criteria, by the requirement to revise feedback, or by self-reflective comparison. An alternative interpretation is that GenAI worked mainly as a trigger for evaluative judgement and internal feedback, while the deeper learning mechanism was reflection and revision (Tai et al., 2018; Nicol, 2021).

The fourth issue concerns feedback accuracy. In the focal study, GenAI evaluated students' feedback comments rather than the original TEL artifacts (Zhou et al., 2026). This means that GenAI may improve the clarity and constructiveness of a comment without verifying whether the comment accurately represents the artifact. A feedback comment can become more detailed but still be pedagogically incomplete, contextually weak, or misaligned with the original work. This concern is important because GenAI systems may generate plausible yet inaccurate or irrelevant feedback (Feldman-Maggor et al., 2025; Jukiewicz, 2024; Usher, 2025).

The fifth issue concerns the social and relational dimensions of feedback. The study found significant improvement in cognitive feedback but not in affective or metacognitive feedback (Zhou et al., 2026). This suggests that GenAI may be more effective in improving the informational quality of feedback than in strengthening dialogic, relational, or reflective aspects of peer assessment. In teacher education, feedback is not only a technical act of pointing out problems and proposing solutions. It is also a professional practice involving trust, empathy, responsibility, and sensitivity to learners (Cheng et al., 2023; Shao et al., 2024; Zeng & Ravindran, 2025).

The sixth issue concerns the quasi-experimental design. Although the study involved experimental and control groups, the use of intact classes means that unmeasured differences may remain (Zhou et al., 2026). Class culture, prior ability, comfort with GenAI, peer interaction norms, and writing competence may have influenced the results. The authors also noted that clustering effects were not included in the statistical analysis. This matters because students were nested in classes, and class-level variation may affect the precision of treatment estimates.

A cautious conclusion is therefore needed. The most defensible interpretation is that GenAI-supported peer feedback was associated with improved feedback quality, stronger feedback uptake, better TEL artifact revision, and modest improvement in self-reported critical thinking tendency among preservice teachers (Zhou et al., 2026). However, the study does not yet prove that GenAI directly causes durable critical thinking performance. A stronger interpretation is that GenAI can support the conditions for critical thinking when it is embedded in a structured environment that promotes feedback literacy, evaluative judgement, self-feedback, and reflective revision.

IMPLICATIONS FOR RESEARCH AND PRACTICE

Future research should examine the mechanisms behind GenAI-supported peer feedback more directly. A useful design would compare several conditions: traditional peer feedback, rubric-supported peer feedback, GenAI-supported feedback without required revision, GenAI-supported feedback with required revision, and GenAI-supported feedback with explicit self-reflection prompts. Such a design would help clarify whether the key mechanism is AI feedback, feedback revision, self-feedback uptake, evaluative judgement, or their combination.

Future studies should also use more direct measures of critical thinking. Self-report questionnaires can capture students’ perceived tendencies, but they should be complemented with performance-based assessments. In teacher education, such assessments could include evaluating flawed lesson plans, critiquing AI-generated instructional suggestions, analyzing student misconceptions, designing technology-enhanced learning activities, or justifying revisions to teaching artifacts. These tasks would provide stronger evidence of whether GenAI-supported peer feedback improves actual reasoning.

Longitudinal research is also needed. If the goal is to develop reflective teachers, researchers should examine whether students continue to apply evaluative criteria after the intervention ends. Follow-up studies could investigate whether preservice teachers transfer feedback literacy and self-reflection into microteaching, lesson planning, classroom observation, or teaching practicum contexts.

For teacher education practice, the focal study suggests that GenAI should be integrated into peer assessment through structured pedagogical design. Students need rubrics, feedback criteria, revision expectations, and opportunities to compare their original feedback with AI-supported revisions. They should also be asked to explain why they accept, reject, or modify GenAI suggestions. This is important because critical thinking requires judgment, not compliance with automated advice.

Teacher educators should also make self-feedback an explicit part of peer assessment. Students can be prompted to ask whether the weakness they identified in a peer's artifact also appears in their own work. They can also be asked to identify which evaluative criteria they used for peers and how those criteria apply to their own design. In this way, peer assessment becomes a reflective loop rather than a one-way exchange of comments.

Finally, GenAI-supported peer feedback should be framed as a tool for developing feedback literacy and professional judgment. GenAI may help students produce clearer and more actionable feedback, but it cannot replace the teacher's responsibility to reason, evaluate context, and make ethical pedagogical decisions. In preservice teacher education, the ultimate goal is not better AI-assisted comments, but more reflective and critically aware future teachers.

CONCLUSION

The focal study provides important evidence that GenAI-supported peer feedback can improve the quality of peer assessment in teacher education. Its strongest contribution lies in showing that GenAI can serve as a feedback-quality scaffold within a structured peer assessment process. The study also highlights the importance of self-feedback uptake, suggesting that the benefits of GenAI-supported peer feedback for critical thinking depend on whether students apply evaluative criteria to their own work.

At the same time, the findings should be interpreted carefully. The study demonstrates clear gains in feedback quality, feedback uptake, and TEL artifact improvement, but the evidence for critical thinking is based on self-reported tendency and shows a smaller effect. The intervention was also relatively short, and the specific contribution of GenAI cannot be fully separated from rubric use, revision practice, self-feedback application, and evaluative judgement.

The key lesson is that better feedback is not enough. For GenAI-supported peer assessment to foster critical thinking, students must engage in self-reflection, apply feedback meaningfully, and develop the capacity to judge both human and AI-generated suggestions. In teacher education, GenAI should therefore be designed not as a shortcut to better comments, but as a scaffold for reflective professional judgment.

Author Contributions

The authors have sufficiently contributed to the study, and have read and agreed to the published version of the manuscript.

Funding

This commentary received no external funding.

Acknowledgement

The author would like to acknowledge the scholarly contribution of the empirical study discussed in this commentary, which provided the basis for critical reflection on GenAI-supported peer assessment in teacher education.

Declaration of Interest

The author declares no conflict of interest.

REFERENCES

Abrami, P. C., Bernard, R. M., Borokhovski, E., Surkes, M. A., Wade, A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78*(4), 1102–1134. https://doi.org/10.3102/0034654308326084

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85*(2), 275–314. https://doi.org/10.3102/0034654314551063

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education, 21*(1), 23. https://doi.org/10.1186/s41239-024-00455-4

Banihashem, S. K., Noroozi, O., Van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review, 37*, 100489. https://doi.org/10.1016/j.edurev.2022.100489

Barahona, C., Nussbaum, M., Martin, V., Meneses, A., Arriagada, S., Di Serio, A., & Hilliger, I. (2023). Technology-scaffolded peer assessment for developing critical thinking in pre-service teacher training: The importance of giving feedback. *Educational Technology Research and Development, 71*(2), 667–688. https://doi.org/10.1007/s11423-022-10173-1

Bozkurt, A., Xiao, J., Farrow, R., Bai, J. Y. H., Nerantzi, C., Moore, S., Dron, J., Stracke, C. M., Singh, L., Crompton, H., Koutropoulos, A., Terentev, E., Pazurek, A., Nichols, M., Sidorkin, A. M., Costello, E., Watson, S., Mulligan, D., Honeychurch, S., … Asino, T. I. (2024). The manifesto for teaching and learning in a time of generative AI: A critical collective stance to better navigate the future. *Open Praxis, 16*(4), 487–513. https://doi.org/10.55982/openpraxis.16.4.777

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Chen, W., Tan, J. S. H., & Pi, Z. (2021). The spiral model of collaborative knowledge improvement: An exploratory study of a networked collaborative classroom. *International Journal of Computer-Supported Collaborative Learning, 16*(1), 7–35. https://doi.org/10.1007/s11412-021-09338-6

Cheng, L., Li, Y., Su, Y., & Gao, L. (2023). Effect of regulation scripts for dialogic peer assessment on feedback quality, critical thinking and climate of trust. *Assessment & Evaluation in Higher Education, 48*(4), 451–463. https://doi.org/10.1080/02602938.2022.2092068

Cho, Y. H., & Cho, K. (2011). Peer reviewers learn from giving comments. *Instructional Science, 39*(5), 629–643. https://doi.org/10.1007/s11251-010-9146-1

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi, 37*(1), 165–184. https://doi.org/10.1007/s11245-016-9401-4

Er, E., Dimitriadis, Y., & Gašević, D. (2021). Collaborative peer feedback and learning analytics: Theory-oriented design for supporting class-wide interventions. *Assessment & Evaluation in Higher Education, 46*(2), 169–190. https://doi.org/10.1080/02602938.2020.1764490

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. California Academic Press.

Facione, P. A. (2020). *Critical thinking: What it is and why it counts*. Measured Reasons LLC.

Fang, J.-W., Chang, S.-C., Hwang, G.-J., & Yang, G. (2021). An online collaborative peer-assessment approach to strengthening pre-service teachers' digital content development competence and higher-order thinking tendency. *Educational Technology Research and Development, 69*(2), 1155–1181. https://doi.org/10.1007/s11423-021-09990-7

Feldman-Maggor, Y., Blonder, R., & Alexandron, G. (2025). Perspectives of generative AI in chemistry education within the TPACK framework. *Journal of Science Education and Technology, 34*(1), 1–12. https://doi.org/10.1007/s10956-024-10147-3

Gao, X., Noroozi, O., Gulikers, J., Biemans, H. J. A., & Banihashem, S. K. (2024). A systematic review of the key components of online peer feedback practices in higher education. *Educational Research Review, 42*, 100588. https://doi.org/10.1016/j.edurev.2023.100588

Guo, K. (2024). EvaluMate: Using AI to support students' feedback provision in peer assessment for writing. *Assessing Writing, 61*, 100864. https://doi.org/10.1016/j.asw.2024.100864

Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education, 63*, 100962. https://doi.org/10.1016/j.iheduc.2024.100962

Guo, K., Zhang, E. D., Li, D., & Yu, S. (2025). Using AI-supported peer review to enhance feedback literacy: An investigation of students' revision of feedback on peers' essays. *British Journal of Educational Technology, 56*(4), 1612–1639. https://doi.org/10.1111/bjet.13540

Halpern, D. F. (2014). *Thought and knowledge: An introduction to critical thinking* (5th ed.). Psychology Press. https://doi.org/10.4324/9781315885278

Han, Y., & Xu, Y. (2020). The development of student feedback literacy: The influences of teacher feedback on peer feedback. *Assessment & Evaluation in Higher Education, 45*(5), 680–696. https://doi.org/10.1080/02602938.2019.1689545

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Jukiewicz, M. (2024). The future of grading programming assignments in education: The role of ChatGPT in automating the assessment and feedback process. *Thinking Skills and Creativity, 52*, 101522. https://doi.org/10.1016/j.tsc.2024.101522

Kerman, N. T., Banihashem, S. K., Karami, M., Er, E., Van Ginkel, S., & Noroozi, O. (2024). Online peer feedback in higher education: A synthesis of the literature. *Education and Information Technologies, 29*(1), 763–813. https://doi.org/10.1007/s10639-023-12273-8

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193–211. https://doi.org/10.1080/02602938.2019.1620679

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525–536. https://doi.org/10.1111/j.1467-8535.2009.00968.x

Liu, J., McBride, R. E., Xiang, P., & Scarmardo-Rhodes, M. (2018). Physical education pre-service teachers' understanding, application, and development of critical thinking. *Quest, 70*(1), 12–27. https://doi.org/10.1080/00336297.2017.1330218

Lu, G., & Ba, S. (2025). Exploring the impact of GAI-assisted feedback on pre-service teachers' situational engagement and performance in inquiry-based online discussion. *Educational Psychology*, 1–26. https://doi.org/10.1080/01443410.2025.2489784

Lu, J., Zheng, R., Gong, Z., & Xu, H. (2024). Supporting teachers' professional development with generative AI: The effects on higher order thinking and self-efficacy. *IEEE Transactions on Learning Technologies, 17*, 1267–1277. https://doi.org/10.1109/TLT.2024.3369690

Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education, 46*(5), 756–778. https://doi.org/10.1080/02602938.2020.1823314

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational Research Review, 9*, 114–128. https://doi.org/10.1016/j.edurev.2012.12.002

Noroozi, O., Alqassab, M., Taghizadeh Kerman, N., Banihashem, S. K., & Panadero, E. (2025a). Does perception mean learning? Insights from an online peer feedback setting. *Assessment & Evaluation in Higher Education, 50*(1), 83–97. https://doi.org/10.1080/02602938.2024.2345669

Noroozi, O., Haddadian, G., Gao, X., Schunn, C., Alqassab, M., & Banihashem, S. K. (2025b). The value of GenAI for peer feedback provision: Student perceptions and impacts. *International Journal of Educational Technology in Higher Education, 22*(1), 61. https://doi.org/10.1186/s41239-025-00558-6

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/BF00117714

Schmid, M., Brianza, E., & Petko, D. (2021). Self-reported technological pedagogical content knowledge (TPACK) of pre-service teachers in relation to digital technology use in lesson plans. *Computers in Human Behavior, 115*, 106586. https://doi.org/10.1016/j.chb.2020.106586

Shao, J., Cheng, L., Wang, Y., Li, K., & Li, Y. (2024). How peer feedback with regulation scripts contributes to the development of critical thinking in dialogues: Strengthening cognitive and affective feedback content. *Interactive Learning Environments, 32*(10), 6121–6140. https://doi.org/10.1080/10494820.2023.2251040

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work of self and others. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

Topping, K. J. (2009). Peer assessment. *Theory Into Practice, 48*(1), 20–27. https://doi.org/10.1080/00405840802577569

Usher, M. (2025). Generative AI vs. instructor vs. peer assessments: A comparison of grading and feedback in higher education. *Assessment & Evaluation in Higher Education*, 1–16. https://doi.org/10.1080/02602938.2025.2487495

Valero Haro, A., Noroozi, O., Biemans, H. J. A., Mulder, M., & Banihashem, S. K. (2024). How does the type of online peer feedback influence feedback quality, argumentative essay writing quality, and domain-specific learning? *Interactive Learning Environments, 32*(9), 5459–5478. https://doi.org/10.1080/10494820.2023.2215822

van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R.-J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review, 20*, 24–34. https://doi.org/10.1016/j.edurev.2016.10.003

Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wood, J. (2022). Making peer feedback work: The contribution of technology-mediated dialogic peer feedback to feedback uptake and literacy. *Assessment & Evaluation in Higher Education, 47*(3), 327–346. https://doi.org/10.1080/02602938.2021.1914544

Wu, Y., & Schunn, C. D. (2020a). From feedback to revisions: Effects of feedback features and perceptions. *Contemporary Educational Psychology, 60*, 101826. https://doi.org/10.1016/j.cedpsych.2019.101826

Wu, Y., & Schunn, C. D. (2020b). When peers agree, do students listen? The central role of feedback quality and feedback frequency in determining uptake of feedback. *Contemporary Educational Psychology, 62*, 101897. https://doi.org/10.1016/j.cedpsych.2020.101897

Yeh, Y.-F., Chan, K. K. H., & Hsu, Y.-S. (2021). Toward a framework that connects individual TPACK and collective TPACK: A systematic review of TPACK studies investigating teacher collaborative discourse in the learning by design process. *Computers & Education, 171*, 104238. https://doi.org/10.1016/j.compedu.2021.104238

Zeng, X., & Ravindran, L. (2025). Design, implementation, and evaluation of peer feedback to develop students' critical thinking: A systematic review from 2010 to 2023. *Thinking Skills and Creativity, 55*, 101691. https://doi.org/10.1016/j.tsc.2024.101691

Zhan, Y., Boud, D., Dawson, P., & Yan, Z. (2025). Generative artificial intelligence as an enabler of student feedback engagement: A framework. *Higher Education Research & Development, 44*(5), 1289–1304. https://doi.org/10.1080/07294360.2025.2476513

Zhan, Y., & Yan, Z. (2025). Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1–14. https://doi.org/10.1080/02602938.2025.2471821

Zhan, Y., Yan, Z., Wan, Z. H., Wang, X., Zeng, Y., Yang, M., & Yang, L. (2023). Effects of online peer assessment on higher-order thinking: A meta-analysis. *British Journal of Educational Technology, 54*(4), 817–835. https://doi.org/10.1111/bjet.13310

Zhang, Y., Chen, H., Pi, Z., & Yang, J. (2024). Interactive equality in peer assessment: The impacts on preservice teachers' technology-enhanced learning design and feedback uptake. *Teaching and Teacher Education, 138*, 104408. https://doi.org/10.1016/j.tate.2023.104408

Zhang, Y., Chen, H., Wang, J., Pi, Z., & Yang, J. (2025). The impacts of self-monitoring prompt on provision and uptake of peer feedback and on revision performance: A mixed methods study. *Studies in Educational Evaluation, 84*, 101442. https://doi.org/10.1016/j.stueduc.2024.101442

Zhang, Y., Pi, Z., Chen, L., Zhang, X., & Yang, J. (2021). Online peer assessment improves learners' creativity: Not only learners' roles as an assessor or assessee, but also their behavioral sequence matter. *Thinking Skills and Creativity, 42*, 100950. https://doi.org/10.1016/j.tsc.2021.100950

Zhou, J., Zhao, N., Gao, X., & Han, X. (2026). The effects of generative AI-supported peer feedback on preservice teachers' learning outcomes, feedback uptake, and critical thinking. *Thinking Skills and Creativity, 62*, 102236. https://doi.org/10.1016/j.tsc.2026.102236