

From Teacher Expertise to Feedback Design: A Module and Preliminary Guidelines for Unplugged ITS

Laíza Ribeiro^{1,2}, Camila Fior³, Luiz Rodrigues^{4,5}, Diego Dermeval⁵, Seiji Isotani⁶

¹Institute of Technology and Leadership (Inteli)

²Institute of Mathematics and Computer Sciences - University of São Paulo (USP)

³Faculty of Education - State University of Campinas (UNICAMP)

⁴Federal Technological University of Paraná (UTFPR)

⁵Center for Excellence in Social Technologies - Federal University of Alagoas (NEES)

⁶Graduate School of Education - University of Pennsylvania (UPenn)

laiza@inteli.edu.br, cafior@unicamp.br, luiz.rodrigues@nees.ufal.br

diego.matos@famed.ufal.br, sisotani@upenn.edu

Abstract. *Intelligent Tutoring Systems (ITS), traditional or unplugged, aim to foster personalized learning, often relying on feedback mechanisms. However, many approaches focus on automated feedback without explicitly addressing the pedagogical criteria underlying its design or teacher involvement in this process. This study contributes to the field of Artificial Intelligence in Education by proposing (i) a Feedback Module and (ii) a set of preliminary guidelines for feedback design, with a focus on Unplugged ITS, explicitly incorporating teachers as active participants in the feedback design process. Using a mixed-methods approach, we conducted interviews with 14 elementary school teachers, analyzed through Bardin's Content Analysis, to derive the Feedback Module, and validated the AI-generated feedback through a Likert-scale survey with the same teachers. An initial evaluation of the generated feedback found positive results, particularly in clarity, motivational language, and support for mathematical reasoning.*

1. Introduction

Feedback is central to learning and formative assessment, enhancing knowledge acquisition, self-regulation, and teacher-student engagement [Hattie and Timperley 2007, Deeva et al. 2021, Wisniewski et al. 2020, Boud and Molloy 2013].

Technologies such as Intelligent Tutoring Systems (ITSs) are often developed using Artificial Intelligence (AI) and applied to support feedback in educational contexts. ITSs aim to support students by providing individualized guidance and tailored feedback throughout the learning process [Sleeman and Brown 1982, Soofi and Ahmed 2019], offering benefits such as personalized learning experiences, adaptive feedback, and increased learner autonomy [Ma et al. 2014, Xu et al. 2019]. However, implementing this technology requires adequate infrastructure, which is not always available in many educational settings. Unplugged ITSs have emerged as a promising alternative for resource-

constrained environments, enabling students to experience the pedagogical benefits of intelligent systems without requiring robust technological infrastructure such as individual access to computers or internet connectivity [Isotani et al. 2023]. This approach creates new opportunities for embedding pedagogically grounded feedback practices in diverse, AI-powered educational settings [Rodrigues et al. 2024].

It is important to clarify what “unplugged” means in this context: the student interacts with the ITS through physical materials (e.g., printed worksheets, cards, or paper-based activities) and does not require individual access to a device or internet connection. The teacher, acting as the system’s mediator, uses a connected device to process the student’s response through the proposed Feedback Module, which integrates a GenAI model as its core computational engine, and then delivers the resulting feedback to the student in a non-digital format (e.g., written or verbal). This design is consistent with the Unplugged ITS paradigm [Isotani et al. 2023], in which AI-driven intelligence operates at the teacher’s level rather than requiring student-facing technology.

Despite documented advances and benefits associated with ITSs, a gap remains in the systematic design of feedback, specifically, how feedback is designed, contextualized, and aligned with instructional intentions. A review of the literature by [Silva et al. 2025] reveals the absence of structured models, frameworks, or processes that explicitly incorporate teachers’ pedagogical expertise and integrate human and AI to support effective student learning in ITSs.

This limitation extends beyond ITS research. Although areas such as Human-AI Collaboration, Teacher-in-the-Loop, and Explainable AI increasingly recognize teachers as essential stakeholders in AI-supported education [Lu et al. 2026, Khosravi et al. 2022, Holstein et al. 2019], existing work has primarily explored teachers’ roles in interacting with, supervising, or interpreting AI systems, while comparatively less attention has been devoted to involving teachers in the pedagogical design of feedback mechanisms.

To address that gap, this paper proposes a Feedback Module and a set of preliminary guidelines for mathematics feedback in Unplugged ITSs, developed through interviews with elementary school teachers. The proposed module incorporates distinct structures for feedback on correct and incorrect responses. Additionally, we validate this module with elementary education teachers, leveraging recent advances in GenAI to support the module’s operationalization while constraining them through pedagogically grounded design principles. The module is designed to be integrated into an Unplugged ITS to support teachers as active designers of pedagogically sound feedback. Accordingly, besides proposing the Feedback Module, this study addresses the following Research Question:

- **RQ:** How do teachers perceive the pedagogical quality of AI-generated mathematics feedback, and how can these perceptions inform the design of feedback in Unplugged ITSs?

2. Related Work

This section reviews studies on both Traditional and Unplugged ITSs, focusing on the role attributed to teachers in the feedback design process.

[Arends et al. 2017] and [Viswanathan et al. 2020] investigate automated feedback mechanisms in Traditional ITSs and report positive effects on students’ problem-

solving performance and learning outcomes. [Arends et al. 2017] present the MicK Tutor, which guides students through evaluating microcontroller I/O expressions using a rule-based domain reasoner implemented in the Ideas framework. Although instructors can customize exercises and feedback messages during the system configuration stage, they do not actively participate in pedagogical design or in the generation of feedback during operation. Similarly, [Viswanathan et al. 2020] propose *Mechanix*, an ITS that provides automated feedback for free-body diagrams through sketch recognition techniques. In this approach, feedback is generated entirely through pattern recognition without teacher intervention or customization. Consequently, both studies assign teachers only indirect roles, such as configuring system parameters or validating instructional content, rather than involving them as active stakeholders in the design and refinement of pedagogical feedback strategies.

These findings are consistent with the results of our recent Systematic Literature Review on feedback in ITSs [Silva et al. 2025], which analyzed 27 empirical studies. The review showed that, although automated feedback consistently improves student performance, teacher participation in feedback design remains limited. Only 10 of the 27 studies involved teachers in any stage of feedback elaboration, and their role was generally restricted to validating messages, configuring system parameters, or supplying predefined content, rather than collaboratively designing pedagogical feedback strategies. Furthermore, the review identified the absence of systematic methods for integrating teachers' pedagogical expertise into feedback design as a major research gap, highlighting the need for approaches that explicitly support teacher participation throughout the feedback design process.

In contrast to Traditional ITS studies conducted in higher education, more recent work has explored Unplugged ITS approaches for educational settings with limited technological infrastructure and closer alignment with everyday teaching practices. Studies such as [Portela et al. 2023] report the deployment of AIED Unplugged at scale for writing instruction, using computer vision and NLP to automatically assess handwritten essays and generate paper-based feedback and dashboards for teachers. However, feedback authoring remained system-driven: teachers received and interpreted AI-generated diagnostics to group and support students, rather than composing the feedback themselves.

Overall, prior studies indicate a lack of a systematic approach to feedback design, particularly regarding the integration of teachers and their pedagogical expertise, often resulting in ad hoc or system-driven feedback practices. Consequently, automated feedback may not be designed with the consistency and pedagogical quality required to effectively support students' learning, as teachers are rarely involved in its conception.

3. Method

This research is a mixed-methods exploratory study, conducted in two steps. In the first step, interviews were conducted and analyzed using Bardin's Content Analysis approach [Bardin 1977]. In the second step, the generated feedback was validated through a survey combining Likert-scale items and open-ended questions, analyzed using statistical tests.

Each interview was conducted individually with 14 elementary school teachers (early grades), lasting approximately 30 to 45 minutes, and followed a protocol designed to elicit teachers' feedback practices, pedagogical criteria, and expectations regarding

student responses. Transcripts underwent Bardin's Content Analysis across its three standard phases (pre-analysis, material exploration, and results treatment). The inter-rater agreement between two independent coders yielded a Cohen's Kappa of 0.885, indicating near-perfect agreement. The resulting Feedback Module was subsequently evaluated by the same teachers through a survey comprising 10 AI-generated feedback instances, as described in Section 3.2.

3.1. Feedback Module

Based on the Content Analysis described above, we identified distinct structures for correct and incorrect feedback, grounded in teachers' actual feedback practices. The resulting structure is shown in Table 1.

Table 1. Structure of Feedback

Incorrect Feedback	Correct Feedback
Initial Acknowledgment	Initial Acknowledgment
Step-by-Step Process	Specific Mention of Success or Progress
Guidance and Reflection for Improvement	Guidance and Reflection for Improvement

The opening and closing steps are identical for both correct and incorrect feedback, with only the second step differing. A detailed explanation of each step is provided in the document *Structure of Feedback.pdf*, available at <https://bit.ly/sbie-2026>.

Based on these steps, we proposed a module to organize the feedback generation process:

- **Input:** The Unplugged ITS receives the student's exercise answer.
- **Processing:** In this phase, we use the Gemini 3.1 Pro¹ to analyze the student's exercise answer, and to write feedback using the previously presented structure. The prompt is presented below:
"Create feedback for a mathematics exercise in an accessible language for children in the early years of elementary school. The feedback should follow this structure: if the answer is incorrect, follow the steps: (i) initial recognition, (ii) step-by-step guidance, and (iii) guidance or reflection for improvement. If the answer is correct, follow the steps: (i) initial recognition, (ii) specific mention of the correct answer or progress, and (iii) guidance or reflection for improvement. The feedback should be short (no more than 5 lines) and must not reveal the correct answer."
- **Output:** The teacher receives, reviews, can modify/customize, and forwards the feedback.

3.2. Initial Evaluation

As an initial evaluation of the proposed module, this study focused on examining the feedback generated by Gemini 3.1 Pro using the defined prompt. The goal of this evaluation is to analyze teachers' perceptions regarding the structure, clarity, and pedagogical adequacy of the generated feedback.

¹<https://gemini.google.com/>

Data were collected using a survey instrument validated by a user experience expert with a PhD in the field, who reviewed the instrument once and suggested adjustments to item wording, layout, and usability. The survey comprises 10 student exercise responses, each accompanied by feedback generated by Gemini 3.1 Pro. The survey comprised both five-point Likert scale items and open-ended text fields, allowing teachers to provide qualitative feedback and justify their ratings. For each exercise, the participating teachers evaluated both the student response and the corresponding feedback. The feedback validation criteria were grounded in the literature and informed by interviews with teachers who work with students in the early years of elementary school. Two criteria are shared by feedback on both incorrect and correct exercises, in addition to criteria specific to each type. These criteria are presented in Table 2 and described in greater detail in the supplementary material, in the document "Criteria for Correct and Incorrect Feedback.pdf", available at <https://bit.ly/sbie-2026>.

Table 2. Criteria for Correct and Incorrect Feedback

Correct Feedback	Incorrect Feedback
Is the feedback structured in a clear, logical, and comprehensible manner? ¹	Is the feedback structured in a clear, logical, and comprehensible manner? ¹
Does the feedback specifically acknowledge and value the student's effort and correct responses? ²	Does the feedback provide support that enables the student to carry out the required logical reasoning? ³
Does the feedback promote reflection and learner autonomy in the learning process? ⁴	Does the feedback encourage the student to review their steps? ⁵
Is the language supportive, motivating, and appropriate to the student's level? ⁶	Is the language supportive, motivating, and appropriate to the student's level? ⁶

¹[Hattie and Timperley 2007, Shute 2008] ²[Brookhart 2017]

³[Lipnevich and Smith 2009, Rabbani and Husain 2024]

⁴[Kluger and DeNisi 1996, Nicol and Macfarlane-Dick 2006]

⁵[Hattie and Timperley 2007, Nicol and Macfarlane-Dick 2006]

⁶[Black and Wiliam 1998, Ryan and Deci 2000]

The 10 exercises were selected to cover essential mathematical concepts from the early years of Brazilian elementary education, including addition and subtraction problems presented in a contextual narrative format (word problems). Five exercises were answered correctly by the students and five were answered incorrectly, each with a different type of error (e.g., incorrect selection of the operation, arithmetic error, or partial answer).

Fourteen participants were recruited using purposive sampling. They are teachers currently or previously working with students in the early grades of elementary school. Further information regarding years of experience, school type, and educational background was not collected.

This study was approved by the Ethics Committee through Plataforma Brasil²,

²<https://plataformabrasil.saude.gov.br/login.jsf>

under approval ID CAAE: 73285823.2.0000.5390. All participants provided informed consent prior to their participation.

4. Results

4.1. Correct Exercise Answers and Feedback

Figure 1 presents an example of feedback generated by Gemini 3.1 Pro for correct exercise answers. The full set of feedback texts is available in the file “Feedback for Correct Exercises.pdf”, available at <https://bit.ly/sbie-2026>.

All feedback instances followed the three-part structure summarized in Table 1, with the second step specifically highlighting the student’s success or progress.

Initial Acknowledgment: What a brilliant piece of work! Your reasoning is impeccable.
Specific Mention of Success or Progress: You recognized that João gave some candies away, so you correctly used subtraction to determine how many remained. The final result is correct.
Guidance or Reflection for Improvement: For the next challenge, think about this variation: if João had given 5 candies to his brother, how many candies would he have left?

Figure 1. Example of feedback for correct exercise

As shown in Table 3 (representing 70 evaluations per criterion), teachers predominantly evaluated correct feedback positively. The highest mean scores were observed for supportive language ($M = 4.01$, $SD = 0.79$) and clarity of organization ($M = 3.94$, $SD = 1.05$), with ratings concentrated in the “Agree” and “Strongly Agree” categories. Conversely, the criterion related to promoting reflection and learner autonomy received the lowest mean score ($M = 3.73$, $SD = 1.00$) and more neutral responses, suggesting that teachers view metacognitive step-reviewing as less critical once a student has already achieved a correct resolution.

Table 3. Evaluation of the Feedback for the Correct Exercise Answers

Criteria	1	2	3	4	5	<i>M</i>	<i>SD</i>
Is the feedback structured in a clear, logical, and comprehensible manner?	2	7	7	31	23	3.94	1.05
Does the feedback specifically acknowledge and value the student’s effort and correct responses?	2	8	5	32	23	3.94	1.07
Does the feedback promote reflection and learner autonomy in the learning process?	2	8	9	35	16	3.73	1.00
Is the language supportive, motivating, and appropriate to the student’s level?	0	3	12	36	19	4.01	0.79

Note. Responses rated on a five-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). *M* = Mean; *SD* = Standard Deviation.

4.2. Incorrect Exercise Answers and Feedback

Figure 2 presents an example of feedback generated by Gemini 3.1 Pro for incorrect exercise answers. The full set of feedback texts is provided in the document "Feedback for Incorrect Exercises.pdf", available at <https://bit.ly/sbie-2026>.

All feedback instances followed the three-part structure summarized in Table 1, with the second step providing step-by-step guidance.

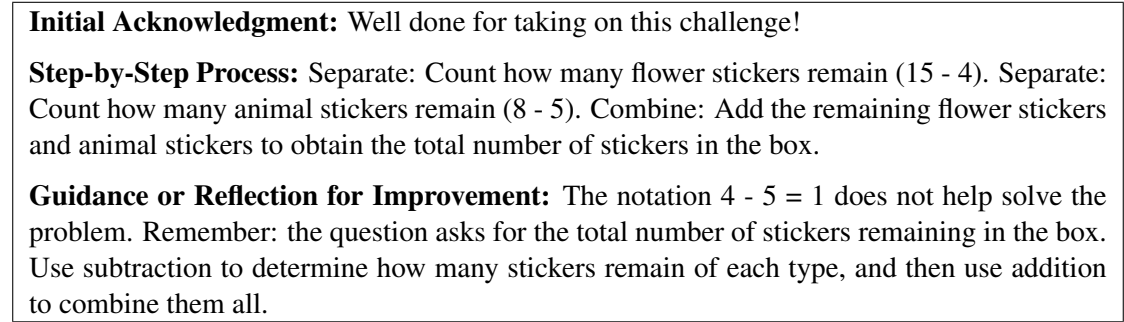


Figure 2. Example of feedback for incorrect exercises

Table 4 reports the frequencies, totaling 70 responses per criterion. Overall, mean scores were lower than those observed for correct feedback, with clarity receiving the highest rating ($M = 3.53$, $SD = 0.97$) and encouragement to review steps the lowest ($M = 3.44$, $SD = 1.00$). Higher levels of agreement were observed for clarity and supportive language ($M = 3.69$, $SD = 0.88$), while criteria associated with support for logical reasoning ($M = 3.56$, $SD = 1.05$) and step-review showed more dispersed responses, with a higher proportion of neutral and disagreement ratings. This indicates greater variability in teachers' evaluations of these aspects, highlighting the challenge of generating automated guidance that perfectly aligns with diverse pedagogical strategies for error intervention.

To examine whether differences between feedback types were statistically meaningful, Mann-Whitney U tests were conducted on shared criteria. Results indicated that correct feedback received significantly higher ratings than incorrect feedback for both clarity ($U=3057$, $p=0.006$, $r=0.23$) and supportive language ($U=2948$, $p=0.025$, $r=0.19$), with small-to-moderate effect sizes. Within each feedback type, Kruskal-Wallis tests revealed no significant differences across criteria (correct: $H=1.94$, $p=0.585$; incorrect: $H=1.88$, $p=0.598$), suggesting that teachers' evaluations were internally consistent.

5. Discussion

Qualitative analysis of teachers' open-ended responses deepens these quantitative patterns. For correct solutions, teachers valued feedback extending beyond mere verification by introducing new challenges or requesting reasoning explanations, aligning with patterns observed in the teacher interviews conducted as part of this study, where process-oriented feedback was consistently prioritized. Conversely, for incorrect solutions, teachers noted that automated feedback favored procedural correction over reasoning strategies. They emphasized that addressing *why* a student erred is vital for formative feedback. These insights reinforce the quantitative variability in reasoning support and step-review, highlighting the need for mechanisms that actively elicit student thinking rather than defaulting to procedural correction.

Table 4. Evaluation of the Feedback for the Incorrect Exercise Answers

Criteria	1	2	3	4	5	M	SD
Is the feedback structured in a clear, logical, and comprehensible manner?	2	13	8	40	7	3.53	0.97
Does the feedback provide support that enables the student to carry out the required logical reasoning?	1	17	4	38	10	3.56	1.05
Does the feedback encourage the student to review their steps?	0	17	14	30	9	3.44	1.00
Is the language supportive, motivating, and appropriate to the student's level?	0	8	17	34	11	3.69	0.88

Note. Responses rated on a five-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). M = Mean; SD = Standard Deviation.

While both feedback types received positive evaluations for clarity and motivating language, qualitative analysis revealed distinct pedagogical gaps [Boud and Molloy 2013]. For correct solutions, teachers valued comments extending beyond verification to support conceptual transfer, yet issues like evaluative language, unsupported inferences, and an overemphasis on final outcomes persisted. These limitations ultimately restrict insight into student thinking and weaken metacognitive reflection [Nicol and Macfarlane-Dick 2006, Shute 2008]. Conversely, feedback on incorrect solutions faced critical evaluations due to an overemphasis on algorithmic procedures over reasoning, conflicting with formative feedback principles [Boud and Molloy 2013]. It often relied on unsupported assumptions and inaccessible language. Thus, feedback must shift from procedural correction to treating errors as formative resources [Black and Wiliam 1998, Hattie and Timperley 2007], as automated guidance lacks impact if student thinking is not explicitly elicited.

5.1. Preliminary Guidelines for Unplugged ITSs

Based on the teachers' evaluations and the limitations identified in the generated feedback, we synthesize a set of preliminary guidelines to inform the design of feedback mechanisms in Unplugged ITSs:

- **G1. Prioritize Reasoning over Procedures:** Feedback should explicitly elicit and support students' underlying mathematical reasoning rather than focusing solely on algorithmic correctness or final outcomes. In practice, this means prompts should ask students to explain *why* a step was taken, not only *what* was computed.
- **G2. Avoid Unsupported Inferences:** Automated feedback mechanisms must refrain from making assumptions about students' internal thinking processes when diagnostic data is limited (e.g., when only the final answer is available). Feedback should be grounded in observable evidence from the student's response.
- **G3. Encourage Self-Regulation:** Incorporate questions that encourage students to review their steps and self-regulate, even when answers are correct. For example: "*Could you solve this in a different way?*" or "*What would happen if the numbers changed?*"

- **G4. Frame Error as a Learning Opportunity:** For incorrect answers, feedback should treat the error as a formative resource rather than a failure. This involves acknowledging what the student attempted, identifying the specific point of divergence, and guiding conceptual reconstruction without simply restating the correct procedure.
- **G5. Recognize the emotional aspects of learning:** Feedback should acknowledge that learning involves emotional as well as cognitive processes. The wording should be supportive, respectful, and appropriate to the student's developmental level, reinforcing competence and motivation while avoiding language that discourages or stigmatizes mistakes. In practice, this means using encouraging expressions and maintaining a constructive tone throughout the feedback.

6. Conclusion

We developed a Feedback Module grounded in an in-depth analysis of interviews with elementary school teachers and examined their evaluations of the pedagogical quality of the generated feedback and its implications for feedback design in an Unplugged ITS.

In response to the RQ guiding this study, the results indicate that feedback was perceived as clear and motivating, especially for correct responses. Evaluations were more critical regarding support for logical reasoning and the promotion of step review and autonomy, particularly in feedback on incorrect answers, which tended to favor procedural correction over reasoning strategies. These findings were translated into the five preliminary guidelines presented in Section 5.1, which highlight the need to strengthen feedback practices that foreground students' reasoning, explicitly articulate problem-solving processes, value multiple solution strategies, and frame errors as opportunities for learning. Feedback that integrates recognition, questioning, and conceptual challenge is essential to enhancing the formative role of assessment and supporting students' autonomy, reflection, and mathematical understanding.

Overall, this study clarifies how teachers evaluate the pedagogical quality of AI-generated feedback in mathematics and how such evaluations can inform the design of feedback mechanisms in Unplugged ITSs. The study contributes two main outcomes: (i) a Feedback Module derived from teachers' interviews and (ii) a set of preliminary guidelines for feedback mechanisms in Unplugged ITSs, grounded in teachers' evaluations of pedagogical quality and translated into actionable design directions under conditions of limited learner data, advancing both the design and theoretical understanding of feedback in ITSs.

This study has some limitations. First, a small number of teachers ($n = 14$) were involved in the evaluation; future work should involve a larger and more diverse sample of teachers. Second, the evaluation focused on teachers' perceptions and did not include validation with real students, which is necessary to understand the module's impact on learning, self-regulation, and academic performance. Third, the module was evaluated in a specific context, namely, elementary school mathematics education in Brazil, and its applicability to other subjects, age groups, or cultural settings remains to be investigated.

Acknowledgments

The authors would like to thank Dr. Guilherme Guerino for his support, as well as the teachers who participated in this study.

Declaration on the Use of Artificial Intelligence

The authors declare that ChatGPT-5.5 and Claude Sonnet 4.5 were used exclusively for grammatical and spelling revision and for the stylistic improvement of the text. Additionally, Gemini 3.1 Pro was used to generate the feedback analyzed in this study. The conceptual content, data analysis, and conclusions were entirely developed by the human authors, who assume full responsibility for the final content and originality of this article.

References

- Arends, H., Keuning, H., Heeren, B., and Jeurig, J. (2017). An intelligent tutor to learn the evaluation of microcontroller i/o programming expressions. In *Proceedings of the 17th Koli Calling International Conference on Computing Education Research*, pages 2–9.
- Bardin, L. (1977). *Análise de Conteúdo*. Editora Persona.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1):7–74.
- Boud, D. and Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, 38(6):698–712.
- Brookhart, S. M. (2017). *How to give effective feedback to your students*. ASCD.
- Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., and De Weerd, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, 162:104094.
- Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1):81–112.
- Holstein, K., McLaren, B. M., and Aleven, V. (2019). Designing for complementarity: Teacher and student needs for orchestration support in ai-enhanced classrooms. In *International conference on artificial intelligence in education*, pages 157–171. Springer.
- Isotani, S., Bittencourt, I. I., Chalco, G. C., Dermeval, D., and Mello, R. F. (2023). Aied unplugged: Leapfrogging the digital divide to reach the underserved. In *International Conference on Artificial Intelligence in Education*, pages 772–779, Cham. Springer Nature Switzerland.
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., and Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and education: artificial intelligence*, 3:100074.
- Kluger, A. N. and DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2):254.
- Lipnevich, A. A. and Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4):319.
- Lu, X., Phyllis Ju, K., Dudley, M., Sano, L., and Wang, X. (2026). Ai-mediated feedback improves student revisions: A randomized trial with feedbackwriter in a large undergraduate course. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pages 1–26.

- Ma, W., Adesope, O. O., Nesbit, J. C., and Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4):901.
- Nicol, D. J. and 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.
- Portela, C., Lisbôa, R., Yasojima, K., Cordeiro, T., Silva, A., Dermeval, D., and Isotani, S. (2023). A case study on aided unplugged applied to public policy for learning recovery post-pandemic in brazil. In *International Conference on Artificial Intelligence in Education*, pages 788–796, Cham. Springer Nature Switzerland.
- Rabbani, L. M. and Husain, S. H. (2024). Fostering student engagement with criticism feedback: importance, contrasting perspectives and key provisions. In *Frontiers in Education*, volume 9, page 1344997. Frontiers Media SA.
- Rodrigues, L., Pereira, F. D., Marinho, M., Macario, V., Bittencourt, I. I., Isotani, S., and Mello, R. (2024). Mathematics intelligent tutoring systems with handwritten input: a scoping review. *Education and Information Technologies*, 29(9):11183–11209.
- Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1):68.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1):153–189.
- Silva, L. R., Fior, C., Rodrigues, L., Penha, R., Dermeval, D., and Isotani, S. (2025). Use of feedback in intelligent tutoring systems: a systematic literature review. *Interactive Learning Environments*, pages 1–22.
- Sleeman, D. and Brown, J. S. (1982). *Intelligent tutoring systems*. Academic Press, London.
- Soofi, A. A. and Ahmed, M. U. (2019). A systematic review of domains, techniques, delivery modes and validation methods for intelligent tutoring systems. *International Journal of Advanced Computer Science and Applications*, 10(3).
- Viswanathan, V. K., Hurt, J. T., Linsey, J. S., Hammond, T. A., Caldwell, B. W., and Talley, K. G. (2020). Impact of a sketch-based tutoring system at multiple universities. In *2020 ASEE Virtual Annual Conference Content Access*.
- Wisniewski, B., Zierer, K., and Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10:487662.
- Xu, Z., Wijekumar, K., Ramirez, G., Hu, X., and Irey, R. (2019). The effectiveness of intelligent tutoring systems on k–12 students’ reading comprehension: A meta-analysis. *British Journal of Educational Technology*, 50(6):3119–3137.