Introducing Generative Feedback to Chatbots in Digital Higher Education

Research in Progress

Dennis Benner

University of Kassel, Information Systems, Kassel, Germany
benner@uni-kassel.de

Abstract. The rapid evolution of the educational landscape, accentuated by the rise of Generative Artificial Intelligence (GAI), calls for a re-evaluation of digital education methodologies to harness new technologies and meet the changing needs of learners. Focusing on the potential of GAI, this study introduces a novel approach called generative feedback and explores its impact on the motivation, learning experience and performance of students in higher education. Therefore, this study presents a prototype that implements a GAI-driven chat using OpenAI's GPT-4 turbo model to provide feedback to students. Using an overlapping study design that combines a longitudinal field study and a two-stage online experiment, this study investigates how generative feedback affects learners. Initial findings suggest that GAI-driven chatbots can provide meaningful feedback to students and enhance their learning experience, setting the stage for further investigation towards a better theoretical understanding of the design and practical application of GAI-driven chatbot feedback.

Keywords: generative artificial intelligence, digital learning, chatbots, generative feedback, higher education.

1 Introduction and Background

The educational landscape is undergoing significant transformation, driven by the emergence of generative artificial intelligence (GAI) and an evolution away from traditional settings towards more individualized settings. Acknowledging ongoing changes, the Organisation for Economic Co-operation and Development have highlighted the urgency of adopting innovative educational practices to meet diverse learning needs (OECD, 2016). This development poses new challenges for digital education to respond to the changes in technology and the diverse needs of learners (Jansen *et al.*, 2022; Tur *et al.*, 2022). Particularly GAI offers strong potentials that need to be tapped for digital education (Huber *et al.*, 2024). Researchers have called to explore this evolution towards more personalized or interactive approaches to digital learning and education (Gupta *et al.*, 2019; Kasneci *et al.*, 2023). In this regard, recent research has focused on the potential of conversational interfaces, such as chatbots, to increase learner engagement and motivation, as well as to provide a learning experiences that is more tailored towards the individual learner (Benner, et al. 2024; Yin *et al.*, 2021; Gupta and Chen, 2022). Studies have also explored collaborative and co-creative learning environments, and the role of feedback in motivating and engaging learners (Bittner *et al.*, 2019; Oeste-Reiß *et al.*, 2016; Wambsganss *et al.*, 2022), as well as self-explanation in learning (Bisra *et al.*, 2018; Hefter and Berthold, 2020). First and foremost feedback has proven to be a valuable tool in various educational settings, particularly adaptive feedback for formative settings e.g., formative assessment (Llorens *et al.*, 2016; Llorens *et al.*, 2014). While there are various forms of delivering feedback to learners, formative feedback has been proven to be highly influential in learning as it can reduce cognitive load, increase motivation, and improve performance for learners (Shute, 2008; Gikandi *et al.*, 2011; Timmers *et al.*, 2013). Formative feedback in itself can range from simple verification (i.e., right or wrong, or a percentage based score) to elaborated feedback which provides an explanation for why a response is true or false while it may or may not present the correct answer (Shute, 2008; Wang *et al.*, 2019). Overarching these extremes, informative tutoring can go beyond that range and provide a combination of different expressions of formative feedback (Narciss and Huth, 2004), which can appeal to a wider scope of learners who may have different perceptions of feedback (Daniels and Bulut, 2020; Say *et al.*, 2023). Prior studies have demonstrated the effectiveness of different types of formative and particularly elaborated feedback in digital learning but also stressed the room and need for further improvements in the age of digital learning (e.g., Zhu *et al.*, 2020; Botelho *et al.*, 2023; Ifenthaler *et al.*, 2023).

Here, GAI has introduced new possibilities and can bear enormous potential to improve digital learning settings (Morgan Stanley, 2023), for instance by using GAI for automated assessment, providing feedback to learners or improving the learning experience in general (Kuklick *et al.*, 2023; Kasneci *et al.*, 2023; Whalen *et al.*, 2023). Albeit the recognized potentials and since GAI is still fairly novel, research faces challenges when there is no state-of-the-art solution, domain knowledge or widely adapted practice for GAI in digital education and its integration with feedback (Huber *et al.*, 2024). Studies addressing this gap and integrate GAI for feedback remain surprisingly scarce with few exceptions that address highly specific cases such as legal or argumentative writing support for law students (Weber *et al.*, 2023; Wambsganss *et al.*, 2022). However, leveraging GAI to provide students feedback could prove to enhance learner motivation and engagement, increase academic performance, or improve the general learning experience. To fill that gap, this paper proposes a novel approach for integrating GAI in digital higher education (i.e., university/college context) through the concept of generative feedback. Therefore, the research objectives of this study are of explorative nature, to investigate how GAI can be used in innovative learning scenarios in higher education and building on that to highlight how future research can follow.

2 Related Research and Hypotheses Development

Chatbots are typically recognized as sophisticated software entities that are capable of interpreting and responding to human language, employing advances in machine learning and broader artificial intelligence to conduct dialogues which emulate human-like interactions (Maedche *et al.*, 2016; Schuetz and Venkatesh, 2020). Studies on chatbots in education show how they can help enhancing the learning experience or academic outcomes among other factors (Schlegel *et al.*, 2023; Tur *et al.*, 2022). These benefits align seamlessly with the evolving educational paradigm that favours agility and self-regulated learning, a preference increasingly expressed by students (Tur *et al.*, 2022). Chatbots, therefore, have the potential to facilitate autonomous pacing in learning, foster greater independence in managing educational activities, and integrate learners into contemporary pedagogical frameworks (Gupta *et al.*, 2019; Janson *et al.*, 2020). From the perspective of educators, chatbots can therefore free up valuable resources. Likewise, previous studies have indicated that chatbots can enhance problem-solving skills, enable collaborative learning, and increase comprehension of educational content (Tolzin and Janson, 2023; Wambsganss *et al.*, 2022). By following this approach, engagement with learning material can be deepened through collaboration or co-creation tasks focused on learner-generated content (Blau and Shamir-Inbal, 2017). However, these collaborative endeavours traditionally rely heavily on human interaction, posing considerable demands on educational personnel and learners alike (Bittner *et al.*, 2019; Oeste-Reiß *et al.*, 2016). Despite recent advances, gaps remain in our understanding of chatbot-facilitated collaboration and content co-creation in digital learning environments (Blau and Shamir-Inbal, 2017; Kang and Santhanam, 2003).

One pertinent topic is the motivation and engagement of students, especially within changing or volatile educational setting that educators have been continuously confronted with during the last few years (Urhahne and Wijnia, 2023). Helping students with feedback in their learning activities is one approach to motivating learners and facilitate a better learning experience (Fidan and Gencel, 2022). Thus, addressing learner motivation and engagement could be achieved by implementing feedback mechanisms that add educational value (Kuklick *et al.*, 2023). Understanding the nature and types of feedback is crucial in the design of information systems that support learning and improvement (Jensen *et al.*, 2021). Formative feedback is widely recognized as a true and tried form of feedback in education and learning (Narciss and Huth, 2004; Shute, 2008). This type of feedback can range from simple types (i.e., verification or confirmation) to complex types (i.e., elaboration). While verification and confirmation focus on a lower level of feedback and can be presented with simple tools such as a true/false icon, a percentage or scoring representation, elaborated feedback requires a more high level textual explanation that elaborates on the student response (Shute, 2008). In more detail, elaborated feedback provides explanations, hints, or suggestions aimed at deepening understanding and facilitating the learner's ability to transfer knowledge to new problems, which has been a focal point of studies in education and learning. For instance, Wang *et al.* (2019) highlighted the potential of elaborated feedback on the learning experience and motivation of students. Further, Zhu *et al.* (2020) corroborated these findings and emphasized on the potential of elaborated feedback on learning gains within a formative assessment task. Discussing the effectiveness of different feedback formats, Daniels and Bulut (2020) analysed the impact of different feedback types and levels. While feedback has been the focus of research in the past (e.g., Wambsganss *et al.*, 2022; Timmers *et al.*, 2013).

GAI has enabled entirely new potentials for providing feedback to learners that could provide more nuanced feedback to individual learners and responses where other approaches not involving human tutors may fail. Consequently, research should investigate how GAI-driven chatbot feedback affects motivation, performance, and the learning experience of students. Thus, this study investigates the effects of low- and high-level generative feedback (i.e., using GAI to create feedback on the fly). Following Shute (2008) this study therefore investigates the effect of generative correct response (i.e., knowledge of correct response) and generative elaborated feedback (i.e., detailed feedback including explanations beyond verification of correctness and correct response). This study then compares how generative correct response feedback performs against generated feedback first and second how this feedback performs in a simple versus complex learning task. While literature suggests that elaborated feedback should in theory be superior to simple feedback like correct response, literature also suggests differences regarding the application to different learning complexities (Moreno, 2004; Butler *et al.*, 2013). Moreover, with the recent advances in technology (i.e., GAI) this should be questioned and investigated regarding the changed playing field in education. Therefore, I raise the following hypotheses:

H1a/b: Generative elaborated feedback outperforms generative correct response feedback in an a) simple and b) complex learning task.

H2: Generative knowledge of correct response feedback does not lead to different outcome for simple versus complex learning tasks.

H3: Generative elaborated feedback performs better in a complex compared to a simple learning task.

3 Research Approach and Design

Since the feedback is part of the study design there are two treatments: a) knowledge of correct response feedback and b) elaborated feedback. Both types of feedback are created with the same OpenAI function (using the GPT-4-turbo model), prompt and knowledge base (e.g., video script, slides, and related book) to ensure that treatments can be compared for evaluation. On the one hand, a) knowledge of correct response feedback is statically created once per learning content (i.e., video), meaning all student responses to a video receive the same knowledge of response feedback. This knowledge of correct response feedback is therefore represented by a GAI-created short summary of the related learning content (e.g., the current short learning video) and displayed as a single block of text. While on the other hand, b) elaborated feedback is dynamically created per student response, i.e., every unique student response receives unique feedback (see Figure 1). This generative elaborative feedback can inform the learner about the “goodness” of their submission, provide corrections if required, and additional tips the learner may find helpful to improve in the future i.e., create better submissions and thus achieve better learning.

The example in Figure 1 highlights how feedback is generated for a false statement with a bad explanation. In this case, the chatbot responds that the feedback does not fit the submitted content and is in general of poor quality (i.e., informs about the goodness), then continues to explain that assessment and simultaneously providing a correct answer, before finally concluding with helpful tips.

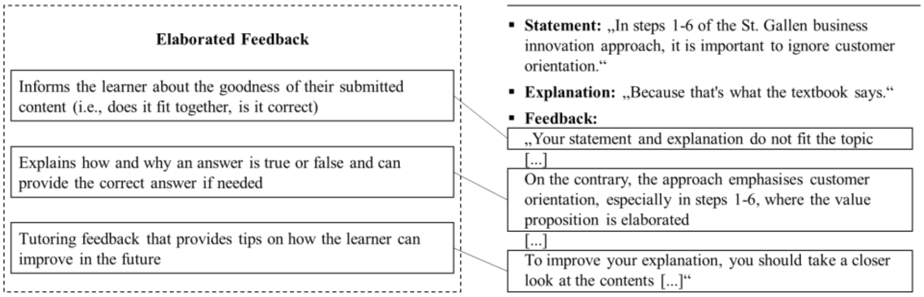

Figure 1. Generative Elaborated Feedback (example excerpt from the field)

To test the hypotheses, I have developed a prototype application for a specific learning task that implements two central components (see architecture in Figure 2). The first component is learning material in form of short learning videos (2-7 minutes) that serve as the basis for the learning task while the second component implements the chatbot which prompts the students for input and delivers the feedback to students. The learning task asks students to watch the short learning videos and create true statements with an additional explanation of the statement.

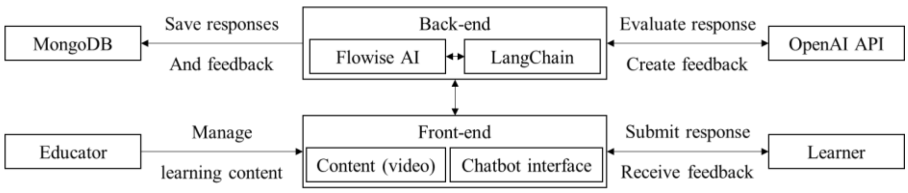

Figure 2. Prototype Architecture

For the evaluation of the prototype and answering the hypotheses, an overlapping study design following Dennis and Valacich (2001) is selected (see Table 1). In the first study the types of feedback will be evaluated against each other i.e., generative correct response versus elaborated feedback, and against a control group with no feedback. The goal is to investigate the influence on a) motivation and b) learning outcome (i.e., the score). In a second follow up study I will then investigate how the feedback design performs in different learning tasks (i.e., simple versus complex). Anderson and Krathwohl (2001) have introduced a revision to Bloom’s original learning taxonomy that forms the foundation for this differentiation where simple tasks refer to the lower and complex to the higher-level learning goal dimensions. Additionally, to the experiment, the generative elaborated feedback is currently undergoing a field study with early bachelor students (i.e., semester 1-3) in an entry level business information systems (i.e., introduction to “Wirtschaftsinformatik”) course at a local university. While not part of the experimental study, this field test will provide insights and potential proof of value for the users (i.e., university students) in the overall context of this research.

Table 1. Overlapping Study Setup
(following Dennis and Valacich, 2001; study 1 = dotted; study 2 = dashed; both studies = solid)

Study 1		
Generative correct response feedback		
Generative elaborated feedback	Simple vs. complex learning task	Study 2
Control Group (no feedback)		

4 Next Steps and Expected Contribution

First, I want to highlight a few first insights from the field. There are currently 57 students actively participating in the ongoing field test who have already contributed over 1100 responses and received feedback. First insights suggest that the artifact is perceived as useful according to the teacher who accompanies the field test. The teacher reports that the general sentiment towards the prototype is positive and that students find it helpful in learning. As I progress with this research, the next steps include carrying out the two-stage online experiment. Unlike the field study that has limitations regarding the sample size, the online experiment will carry a larger sample with a minimum 50 participants per group. In sum, the online experiment will shine light on the effectiveness of the design while the currently ongoing field study will provide insights into the overall value provided to students in higher education. Regardless, there are certain limitations to consider. First, this study focuses only on the extreme ends of the feedback and complexity spectrums. Second, the setting of the experiment may not fit a general experimental platform population as it was designed specifically for students.

Overall, this research specifically investigates the use of an AI-driven chatbot designed to provide generative feedback. Thus, this study aims to make significant contributions in two main areas and the expected contribution is two-fold. Firstly, the study seeks to expand the theoretical framework related to AI-driven educational technologies such as chatbot and AI-feedback design within digital education; particularly to improve motivation, learning experience and academic outcome of students. Secondly, it aims to develop practical design guidelines for educators and enable them to leverage the presented concept for their learning setting. To this end, this study explores the application of GAI in digital education.

Acknowledgements

This research is partially funded by the *Stiftung Innovation in der Hochschule* within the project “Universität Kassel digital: Universitäre Lehre neu gestalten”.

References

Anderson, L. W. and Krathwohl, D. R. (2001), *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, New York: Longman.

Benner, D., Schöbel, S., Janson, A. and Leimeister, J.M. (2024), "Engaging Minds - How Gamified Chatbots can Support and Motivate Learners in Digital Education", *Hawaii International Conference on System Sciences*.

Bisra, K., Liu, Q., Nesbit, J.C., Salimi, F. and Winne, P.H. (2018), "Inducing Self-Explanation: a Meta-Analysis", *Educational Psychology Review*, Vol. 30 No. 3, pp. 703–725.

Bittner, E., Oeste-Reiß, S. and Leimeister, J.M. (2019), "Where is the Bot in our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work", *Hawaii International Conference on System Sciences*, Vol. 52, pp. 284–293.

Blau, I. and Shamir-Inbal, T. (2017), "Re-designed flipped learning model in an academic course: The role of co-creation and co-regulation", *Computers & Education*, Vol. 115, pp. 69–81.

Botelho, A., Baral, S., Erickson, J.A., Benachamardi, P. and Heffernan, N.T. (2023), "Leveraging natural language processing to support automated assessment and feedback for student open responses in mathematics", *Journal of Computer Assisted Learning*, Vol. 39 No. 3, pp. 823–840.

Butler, A.C., Godbole, N. and Marsh, E.J. (2013), "Explanation feedback is better than correct answer feedback for promoting transfer of learning", *Journal of Educational Psychology*, Vol. 105 No. 2, pp. 290–298.

Daniels, L.M. and Bulut, O. (2020), "Students' perceived usefulness of computerized percentage-only vs. descriptive score Reports: Associations with motivation and grades", *Journal of Computer Assisted Learning*, Vol. 36 No. 2, pp. 199–208.

Dennis, A.R. and Valacich, J.S. (2001), "Conducting Experimental Research in Information Systems", *Communications of the Association for Information Systems (CAIS)*, Vol. 7.

Fidan, M. and Gencel, N. (2022), "Supporting the Instructional Videos With Chatbot and Peer Feedback Mechanisms in Online Learning: The Effects on Learning Performance and Intrinsic Motivation", *Journal of Educational Computing Research*, pp. 1–26.

Gikandi, J.W., Morrow, D. and Davis, N.E. (2011), "Online formative assessment in higher education: A review of the literature", *Computers & Education*, Vol. 57 No. 4, pp. 2333–2351.

Gupta, S. and Chen, Y. (2022), "Supporting Inclusive Learning Using Chatbots? A Chatbot-Led Interview Study", *Journal of Information Systems Education*, Vol. 33 No. 1, pp. 98–108.

Gupta, S., Jagannath, K., Aggarwal, N., Sridar, R., Wilde, S. and Chen, Y. (2019), "Artificially Intelligent (AI) Tutors in the Classroom: A Need Assessment Study of Designing Chatbots to Support Student Learning", *Pacific Asia Conference on Information Systems*.

Hefter, M.H. and Berthold, K. (2020), "Preparing learners to self-explain video examples: Text or video introduction?", *Computers in Human Behavior*, Vol. 110.

Huber, S.E., Kiili, K., Nebel, S., Ryan, R.M., Sailer, M. and Ninaus, M. (2024), "Leveraging the Potential of Large Language Models in Education Through Playful and Game-Based Learning", *Educational Psychology Review*, Vol. 36 No. 1.

Ifenthaler, D., Schumacher, C. and Kuzilek, J. (2023), "Investigating students' use of self-assessments in higher education using learning analytics", *Journal of Computer Assisted Learning*, Vol. 39 No. 1, pp. 255–268.

Jansen, R.S., van Leeuwen, A., Janssen, J. and Kester, L. (2022), "Exploring the link between self-regulated learning and learner behaviour in a massive open online course", *Journal of Computer Assisted Learning*, Vol. 38 No. 4, pp. 993–1004.

Janson, A., Söllner, M. and Leimeister, J.M. (2020), "Ladders for Learning: Is Scaffolding the Key to Teaching Problem-Solving in Technology-Mediated Learning Contexts?", *Academy of Management Learning & Education*, Vol. 19 No. 4, pp. 439–468.

Jensen, L.X., Bearman, M. and Boud, D. (2021), "Understanding feedback in online learning – A critical review and metaphor analysis", *Computers & Education*, Vol. 173, p. 104271.

Kang, D. and Santhanam, R. (2003), "A Longitudinal Field Study of Training Practices in a Collaborative Application Environment", *Journal of Management Information Systems*, Vol. 20 No. 3, pp. 257–281.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J. and Kasneci, G. (2023), *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*, EdArXiv.

Kuklick, L., Greiff, S. and Lindner, M.A. (2023), "Computer-based performance feedback: Effects of error message complexity on cognitive, metacognitive, and motivational outcomes", *Computers & Education*, Vol. 200, p. 104785.

Llorens, A.C., Cerdán, R. and Vidal-Abarca, E. (2014), "Adaptive formative feedback to improve strategic search decisions in task-oriented reading", *Journal of Computer Assisted Learning*, Vol. 30 No. 3, pp. 233–251.

Llorens, A.C., Vidal-Abarca, E. and Cerdán, R. (2016), "Formative feedback to transfer self-regulation of task-oriented reading strategies", *Journal of Computer Assisted Learning*, Vol. 32 No. 4, pp. 314–331.

Maedche, A., Morana, S., Schacht, S., Werth, D. and Krumeich, J. (2016), "Advanced User Assistance Systems", *Business & Information Systems Engineering*, Vol. 58 No. 5, pp. 367–370.

Moreno, R. (2004), "Decreasing Cognitive Load for Novice Students: Effects of Explanatory versus Corrective Feedback in Discovery-Based Multimedia", *Instructional Science*, Vol. 32 No. 1/2, pp. 99–113.

Morgan Stanley (2023), "Generative AI in Education: Three Year Outlook", available at: https://www.morganstanley.com/ideas/generative-ai-education-outlook (accessed 5 March 2024).

Narciss, S. and Huth, K. (2004), "How to design informative tutoring feedback for multi-media learning", *Instructional design for multimedia learning*.

OECD (2016), "Better Policies for 2030".

Oeste-Reiß, S., Söllner, M. and Leimeister, J.M. (2016), “Development of a Peer-Creation-Process to Leverage the Power of Collaborative Knowledge Transfer”, *Hawaii International Conference on System Sciences*, pp. 797–806.

Say, R., Visentin, D., Saunders, A., Atherton, I., Carr, A. and King, C. (2023), “Where less is more: Limited feedback in formative online multiple-choice tests improves student self-regulation”, *Journal of Computer Assisted Learning*.

Schlegel, L., Schöbel, S. and Söllner, M. (2023), “Nudging Digital Learning - An Experimental Analysis of Social Nudges to Manage Self-Regulated Learning and Online Learning Success”, *Hawaii International Conference on System Sciences*.

Schuetz, S. and Venkatesh, V. (2020), “The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions of User-System Interaction”, *Journal of the Association for Information Systems*, Vol. 21 No. 2, pp. 460–482.

Shute, V.J. (2008), “Focus on Formative Feedback”, *Review of Educational Research*, Vol. 78 No. 1, pp. 153–189.

Timmers, C.F., Braber-van den Broek, J. and van den Berg, S.M. (2013), “Motivational beliefs, student effort, and feedback behaviour in computer-based formative assessment”, *Computers & Education*, Vol. 60 No. 1, pp. 25–31.

Tolzin, A. and Janson, A. (2023), “Mechanisms of Common Ground in Human-Agent Interaction: A Systematic Review of Conversational Agent Research”, *Hawaii International Conference on System Sciences*.

Tur, G., Castañeda, L., Torres-Kompen, R. and Carpenter, J.P. (2022), “A literature review on self-regulated learning and personal learning environments: features of a close relationship”, *Interactive Learning Environments*, pp. 1–20.

Urhahne, D. and Wijnia, L. (2023), “Theories of Motivation in Education: an Integrative Framework”, *Educational Psychology Review*, Vol. 35 No. 2, pp. 1–35.

Wambsganss, T., Janson, A. and Leimeister, J.M. (2022), “Enhancing argumentative writing with automated feedback and social comparison nudging”, *Computers & Education*, Vol. 191, p. 104644.

Wang, Z., Gong, S.-Y., Xu, S. and Hu, X.-E. (2019), “Elaborated feedback and learning: Examining cognitive and motivational influences”, *Computers & Education*, Vol. 136, pp. 130–140.

Weber, F., Wambsganss, T. and Söllner, M. (2023), “Design and Evaluation of an AI-based Learning System to Foster Students' Structural and Persuasive Writing in Law Courses”, *International Conference on Information Systems*.

Whalen, K.A., Renkl, A., Eitel, A. and Glogger-Frey, I. (2023), “Digital re-attributional feedback in high school mathematics education and its effect on motivation and achievement”, *Journal of Computer Assisted Learning*.

Yin, J., Goh, T.-T., Yang, B. and Xiaobin, Y. (2021), “Conversation Technology With Micro-Learning: The Impact of Chatbot-Based Learning on Students’ Learning Motivation and Performance”, *Journal of Educational Computing Research*, Vol. 59 No. 1, pp. 154–177.

Zhu, M., Liu, O.L. and Lee, H.-S. (2020), “The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing”, *Computers & Education*, Vol. 143, p. 103668.