Didaktik Deutsch

Gloria Goller & Nadine Anskeit

Empirical research on AI-assisted writing

Mixed-methods approaches, long-term effects, and ethical-technological conditions

1 Introduction

With the widespread availability of generative AI tools since the end of 2022, the discussion about writing has changed fundamentally. New possibilities for supporting, reflecting on, and individualizing writing processes are opening up (Kasneci et al., 2023; Rezat & Schindler, 2025), while concerns persist that learners may become dependent on AI and that key writing skills may be weakened (Niloy et al., 2023). Against this backdrop, a central question arises: How can the effects of AI use in writing be researched empirically in a valid and reliable manner?

Existing studies often focus on short-term output analyses, whereas sustainable effects, process data, and motivational dimensions remain underexplored. Therefore, this article aims to systematize key methodological problems in empirical research on AI-assisted writing and to identify potential solutions for intervention studies that address three core requirements: mixed-methods approaches, the examination of long-term effects, and transparent ethical and technological conditions.

As recent work shows (Fütterer et al., 2023; Kasneci et al., 2023), methodological questions cannot be separated from broader educational and ethical considerations, as the use of AI in writing directly affects pedagogical practices, assessment, and notions of authorship. These conditions shape the design, interpretation, and comparability of empirical studies and thus form the framework within which robust research on AI-assisted writing must be developed.

2 Methodological challenges

One central methodological challenge concerns the validity of empirical measurements. AI-supported writing processes can produce text quality that reflects technical assistance rather than actual competence gains (Rezat & Schindler, 2025). Empirical designs must therefore distinguish AI effects from learning processes through the use of, for example, process data, qualitative reflections, or control-group comparisons.

A second challenge is the insufficient integration of process-oriented perspectives. Classical writing research emphasizes product ratings, yet the interaction between writers and AI—described as human–machine "coactivity" (Lehnen & Steinhoff, 2024)—remains underexamined. Log files, prompt developments, chat histories, or self-reports are therefore necessary to capture cognitive and

Didaktik Deutsch
Halbjahresschrift für die Didaktik der deutschen Sprache und Literatur
Special Issue (2026) – Writing with AI. S. 166–170
DOI: 10.21248/dideu.975

metacognitive processes and to complement product analyses, underscoring the need for practice-oriented approaches to AI use in writing that also consider the social dimensions of writing.

A third issue concerns the sustainability of effects. Most studies rely on one-off interventions, although pre–post–follow-up designs are essential for detecting both immediate and delayed learning outcomes (Pissarek & Wild, 2018). This is particularly relevant in AI-assisted writing in which short-term relief may obscure long-term implications for self-regulation and transfer.

Finally, empirical designs must account for ethical and technological conditions. Limited model transparency, data protection issues, and algorithmic biases (Gethmann et al., 2022) influence both research implementation and the interpretation of findings. Transparent reporting of technical parameters and ethical standards is therefore necessary to ensure comparability and replicability.

3 International state of research and methodological designs

International research on writing with AI is characterized by a high degree of methodological and content diversity. The studies discussed here were selected because they represent different methodological orientations (e.g., mixed-methods, quasiexperimental, product-focused, comparative), different writing contexts (EFL, secondary school, creative writing, literary writing, reflective learning settings), and different analytic emphases (output vs. process, short-term vs. long-term effects, human–machine evaluation). Taken together, they illustrate the methodological challenges and opportunities that arise when examining mixed-methods requirements, long-term effects, and ethical and technological conditions.

A prominent example of experimental research is Song and Song's (2023) controlled intervention study on the influence of ChatGPT on the academic writing skills and motivation of EFL students. Their pretest–posttest mixed-methods design, supplemented by qualitative interviews, shows significant improvements in text quality and motivation. Although this triangulation addresses mixed-methods requirements, its short duration also illustrates the need for longitudinal approaches to examine sustainability.

A complementary perspective is offered by Anjarani et al.'s (2024) quasiexperimental study embedded in authentic classroom practice. By focusing on engagement and enjoyment, the study achieves high ecological validity, but offers less experimental control. In comparison to Song and Song, this work highlights a methodological tension between internal validity and ecological validity and between cognitive outcomes and motivational processes.

Product-focused studies such as Bulante et al. (2024) examine AI-assisted and non-AI-assisted writing outputs among secondary school students. Their findings show short-term improvements in linguistic quality, yet without process data or extended time frames, results remain difficult to attribute to actual competence development. Compared to Song and Song or Anjarani et al., Bulante et al.'s study illustrates the limitations of output-only designs and reinforces the need for multimodal data.

The study by Niloy et al. (2023) adds a different perspective by investigating creative writing in a large-scale experiment. AI-assisted texts were more fluent but less original, raising conceptual questions about evaluating creativity under technological influence. Whereas Bulante et al. foreground linguistic quality, Niloy et al. highlight trade-offs between stylistic gains and expressive diversity, thereby introducing an ethical–technological dimension to the necessary evaluation criteria. Führer and Gerjets

(2024) extend this by emphasizing that empirical studies must also account for aesthetic, emotional, and creative dimensions of writing.

From a didactic-methodological standpoint, Rezat and Schindler (2025) argue that empirical research on AI-assisted writing cannot be separated from instructional aims and assessment logics. Their analyses align with approaches that examine the learner perspective, reflective competence, and processual writing development.

A particularly relevant extension of this perspective is found in Graham (2024), who argues that writing remains central in the age of AI because it supports thinking, understanding, and active engagement with content. Accordingly, AI should be integrated in ways that stimulate rather than replace cognitive processes. This theoretical stance frames the design of Goller's (2026) project *Rethinking Writing*, which explicitly investigates how different forms of chatbot interaction can promote writing awareness and argumentative thinking in secondary education.

Goller compares two instructional AI models: a question-guided chatbot that provides prompts and follow-up questions to stimulate deeper reasoning, and a dialogic chatbot that responds to student input without directing it. The study conceptualizes AI not as an optimization tool, but as a didactic catalyst for reflective writing processes. Methodologically, it employs a mixed-methods design: qualitative analyses of chat logs to trace cognitive and metacognitive movements along with quantitative analyses of revised texts focusing on coherence, argumentative structure, and linguistic complexity. Pilot findings (2023) indicate that the question-guided chatbot in particular supports deeper reflective engagement, thereby exemplifying how AI can foster rather than automate thinking. In contrast to product-oriented studies such as Bulante et al., Goller's work foregrounds processual and reflective dimensions, aligning closely with the mixed-methods requirement and ethical–technological considerations regarding the pedagogical purpose of AI use.

Taken together, these studies form an exemplary cross section of international research: They contrast product versus process orientation, short-term versus longitudinal designs, experimental versus practice-based approaches, while also addressing the cognitive, creative, and reflective dimensions of writing. Their methodological differences highlight the need for integrated mixed-methods approaches, sustained perspectives on long-term effects, and transparent ethical–technological reporting standards.

4 Perspectives and outlook

The studies reviewed here indicate that empirical research on AI-assisted writing is still in an early phase of methodological consolidation. The considerable heterogeneity of existing designs suggests that pragmatic, context-dependent approaches have so far prevailed over theoretically coherent methodological frameworks. At the same time, this diversity has produced several developments that are likely to shape the field in the coming years.

First, there is increasing evidence of a paradigm shift toward mixed-methods research that systematically combines quantitative output analyses with qualitative process and reflection data. Such triangulation is essential for generating valid insights into practices of human–machine interaction and for understanding how writing development unfolds in AI-supported settings.

Second, long-term effects are emerging as a key criterion for evaluating the educational value of AI support. Whereas many studies document short-term improvements or motivational gains, robust

empirical knowledge about delayed effects on self-regulation, transfer, and writing awareness remains limited. Longitudinal and follow-up designs are therefore central for future research.

Third, empirical investigations must systematically address ethical and technological conditions. Limited model transparency, algorithmic bias, and data protection constraints affect not only research practice but also the interpretability and comparability of empirical results. Ethical–technological parameters are thus methodologically relevant variables that must be documented explicitly to ensure validity and reproducibility.

Taken together, these developments indicate a conceptual shift away from asking whether AI promotes writing toward examining how, under what conditions, and with what long-term consequences AI systems influence writing processes and outcomes. Advancing empirical research on AI-assisted writing therefore requires an expanded methodological repertoire that integrates mixed-methods logic, long-term perspectives, and transparent ethical and technological reporting. Only through such reflective and context-sensitive approaches can research adequately capture the complexity of interactions between humans, AI systems, and educational environments and contribute to an internationally coherent methodological discourse.

References

Anjarani, S., Furqon, M., & Rakhmawati, F. (2024). *ChatGPT to foster students' engagement in writing class: An intervention study*. *Voices of English Language Education Society, 8*(3), 630–640. https://doi.org/10.29408/veles.v8i3.27687

Bulante, K. M., Morris, D. Q., Jr., Orteza, M. A. T., Balagan, A. M., Taloza, K., & Torres, M. A. (2024). AI-assisted and non-AI-assisted writing performances of Grade 10 students. *Azal Arts and Humanities, 1*(1), 65–94. https://doi.org/10.55687/aah.v1i1.74

Führer, C., & Gerjets, P. (2024). How to understand and write literature with AI? Potenziale und Risiken von KI-Tools für literarisches Lesen und Schreiben. *Medien im Deutschunterricht, 6*(1). https://doi.org/10.18716/OJS/MIDU/2024.1.3

Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in education: Global reactions to AI innovations. *Scientific Reports*, *13*, Article 12018, 1–11. https://doi.org/10.1038/s41598-023-42227-6

Gethmann, C. F., Buxmann, P., Distelrath, J., Humm, B. G., Lingner, S., Nitsch, V., Schmidt, J. C., & Spiecker genannt Döhmann, I. (Eds.). (2022). *Künstliche Intelligenz in der Forschung: Neue Möglichkeiten und Herausforderungen für die Wissenschaft* (Vol. 48). Springer VS. https://doi.org/10.1007/978-3-662-63449-3

Goller, G. (2026). *Schreiben neu denken? Zum Einfluss dialogischer KI-Interaktionen auf das schriftliche Argumentieren von Lernenden der Sekundarstufe I* (Doctoral dissertation in preparation). Department of German Language and Literature, Karlsruhe University of Education.

Graham, S. (2024). Why should we teach writing in the age of artificial intelligence? In *Inspired Ideas*. McGraw Hill. https://medium.com/inspired-ideas-prek-12/why-should-we-teach-writing-in-the-age-of-artificial-intelligence-c49c50300584

Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education.

Learning and Individual Differences, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lehnen, K., & Steinhoff, T. (2024). Digitales Lesen und Schreiben. In J. Androutsopoulos & F. Vogel (Eds.), *Handbuch Sprache und digitale Kommunikation* (pp. 527–546). De Gruyter. https://doi.org/10.1515/9783110744163-025

Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Rahman, S. I. U. (2023). Is ChatGPT a menace for creative writing ability? An experiment. *Journal of Computer Assisted Learning*, *40*(2), 919–930. https://doi.org/10.1111/jcal.12929

Pissarek, M., & Wild, J. (2018). Prä-/Post-/Follow-up-Kontrollgruppendesign. In J. M. Boelmann (Ed.), Empirische Forschung in der Deutschdidaktik (Vol. 1, pp. 215–236). Schneider. https://phbl-opus.phlb.de/frontdoor/deliver/index/docId/590/file/Boelmann_Empirische+Forschung+in+der+Deutschdidaktik_Bd.1_Grundlagen.pdf

Rezat, S., & Schindler, K. (2025). KI-gestützte Textproduktion im Deutschunterricht. *Praxis Deutsch*, *52*(311), 4–11.

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14,* Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Authors' addresses:

Gloria Goller, Karlsruhe University of Education, Bismarckstraße 10, 76133 Karlsruhe, Germany *gloria.goller@gmail.com*

Nadine Anskeit, Karlsruhe University of Education, Bismarckstraße 10, 76133 Karlsruhe, Germany *nadine.anskeit@ph-karlsruhe.de*