

Daniela Matz & Carolin Führer

Methodological approaches to research on writing processes in the context of AI

The example of creative and literary writing

1 Introduction

Under the conditions of generative AI, writing processes have become fundamentally invisible. Unlike traditional writing in which teachers can observe drafting stages, revision marks, and visible struggle, AI-mediated writing collapses the visible evidence of cognitive work. This poses a crucial educational and empirical challenge: Educators and researchers can no longer assume that a finished text reflects the student's actual intellectual labor (Hau, 2025). This is precisely what makes writing-process research so essential, because it makes the invisible visible again. In empirical writing research, the reconstruction of the writing processes themselves is underrepresented compared to the evaluation of the (learners') writing products. In contrast, recent empirical writing-process research (Singh et al., 2023) reveals the need to better understand the relationship between the processes and the products of writing. Especially within the context of AI research, the role of AI in learners' writing development has not yet been studied sufficiently and therefore cannot be assessed accurately. Currently, international writing research is trying to design different AI-support approaches to promote writers' sense of agency and depth of knowledge transformation (Siddiqui et al., 2025). However, in our view, it is imperative to first gain a more profound comprehension of the impact of this technology on the act of writing itself. This understanding is paramount for the establishment of appropriate and effective writing-process support within the context of future writing instruction in educational institutions.

The present article discusses methodological issues in L1 writing-process research by referring to two selected studies. These focus on reconstructing individual (LiSKIDS) and collaborative (WR-AI-TING) writing processes while integrating AI. The former project, LiSKIDS, is funded by the federal state of Baden-Württemberg and explores literary writing.¹ The latter, WR-AI-TING, is funded by the German Federal Ministry of Research, Technology and Space and focuses on creative writing with AI.

After briefly introducing the research questions in each study, the methodological discussion compares and contrasts the two studies under the following rubrics: research questions (2), the writing phases (3), the writing environment and the form of interaction (4), the experimental setup (5), the analysis

¹ "Literary writing" is understood as the process of composing fictional, poetic, or otherwise aesthetically shaped texts. In classroom settings, it is supposed to foster students' awareness of narrative structures, stylistic devices, and the artistic potential of language (Abraham et al., 2023).

and interpretation of the collected data (6), and finally the conclusions (7). The purpose of this comparative presentation is to highlight the methodological issues involved in qualitative research processes in the field of writing.

2 Research questions

The first LiSKIDS study (LiSKIDS 01) focuses on the evaluative processes of writers during AI-assisted literary writing and on the influence of these evaluations on the composition. To address the research questions, the aim is to get as close as possible to the cognitive and emotional processes—that is, to examine them in the very moment of their emergence. LiSKIDS aims at answering the following research questions:

1. How can literary writing processes with and without artificial intelligence be modeled?
2. What influence do literary evaluations exert on the direction and development of writing processes in both settings?
3. Which narrative-related normative values shape the writing processes in both settings?

The central research question guiding WR-AI-TING concerns the manner in which students revise creative texts in co-constructive writing processes both with and without the utilization of AI. The following subquestions are posed:

1. What (peer) interaction practices shape the writing processes with and without AI?
2. What orientational frameworks for creativity shape the writing processes with and without AI?
3. What metareflections on writing become visible in AI-supported and non-AI-supported collaborative settings?

3 Writing phases

Human–machine collaborative writing currently dominates the discussion around meaningful AI applications (Steinhoff, 2025). However, outsourcing too much work and learning to AI systems raises the risk of deskilling (Ferdman, 2025; Führer & Gerjets, 2024). It is particularly the early stages of writing that may be impacted by an intense collaboration with AI, because these stages are characterized by openness: Writers must gather ideas and come up with initial formulations. In order to better understand these dynamics that arise from collaboration with AI in relation to storytelling, LiSKIDS 01 focuses on an early phase of the writing process. It aims to gain a better understanding of the specific effects of human–machine collaboration on storytelling, especially among schoolchildren. What narrative patterns are implemented in such collaborative processes, with which storytelling functions are they associated, and how does human–machine collaboration change the writer’s relationship to storytelling and to their products?

In contrast, WR-AI-TING focuses on later stages of the writing process. Its main emphasis is on revision in creative writing. According to Hayes (2012, p. 376), revisions should not be understood as separate writing processes, but rather as a writing activity that focuses on the writing goal. Here, the function of writing processes as problem-solving processes becomes particularly clear: Revisions are initiated by identifying problems in the existing draft text and then replacing or modifying those text passages that are found to be problematic (Hayes, 2012, p. 376). The reader’s perspective comes to the fore, promoting both text awareness and abstract thinking; and this, in turn, simplifies the formulation of

writing goals. Because inconsistencies in one's own texts are very difficult to detect, it is advisable to involve readers who did not write the text in the revision process (Philipp, 2015, p. 50). In this context, passively accepting AI suggestions without critical evaluation would pose a key risk to the revision process. The authority of AI feedback might tempt students to stop questioning whether the suggestions make sense for their text or whether, conversely, they undermine their original intention. Hence, it is necessary to empower students to evaluate AI results independently. Focusing on learners' discussions about AI feedback in empirical research will help to develop appropriate didactic interventions to counter this temptation, because the metareflection on AI feedback is the decisive learning moment.

4 Writing environment: Interactions and tasks

Writing tasks and the social setting in which they are carried out are central to writing processes. What dynamics shape writing with peers, and how do such environments differ from working alone? What opportunities do learners have to distance themselves from AI output in each of these two settings?

LiSKIDS 01 examines the writing processes of 16 learners in the 9th and 10th grades of intermediate and higher secondary schools (8 from each school type). The writing task involved selecting one of three optional beginnings for short stories that the learners were then asked to continue. Half of the students had the opportunity to use AI and were given general instructions on its use. All students had 50 min to complete the writing assignment. Revisions were not an explicit part of the assignment, but students could make them independently during the writing process. Students worked alone, which is the case in many learning situations, without a teacher, tutor (human or machine), or other students.

The WR-AI-TING study explored the writing discourse between secondary school learners ($n = 118$) who wrote creative texts together in groups with and without AI support. As a first step, learners took part in a brief workshop on Franz Kafka's writing habits, for example, on the characteristically extensive revision processes he performed on his work. Next, they watched a custom-made instructional video about creative writing with AI. In this video, the researchers used thinking aloud to model how AI can be integrated into creative writing processes. Following this, the students had 20 min to write a creative text continuing the first sentence of Kafka's *The Trial*. Finally, the learners were asked to decide which of the two individually created drafts they found to be the more creative, and then to revise it jointly. In addition, the students received structuring tips for revision based on Torrance's (1974) psychological concept of creativity together with more recent, well-regarded models (Sternberg, 2003; Sternberg & Lubart, 1995). These guidelines focused on four key dimensions: fluency (i.e., ability to generate many ideas), flexibility (i.e., ability to produce different ideas), originality (i.e., ability to produce unique and statistically rarer ideas), and elaboration (i.e., ability to elaborate on ideas and develop them in detail). Students were free to decide whether to incorporate these tips into their revision or to ignore them.

5 Experimental setup

Both studies used a writing tool developed specifically for research purposes in collaboration with the Leibniz-Institut für Wissensmedien (IWM) in Tübingen. This AI-integrated tool uses ChatGPT (version 4.0) as its underlying language model and embeds the chat interface directly into the writing environment. Whereas the students' writing conversations (WR-AI-TING) or their think-aloud verbalizations (LiSKIDS 01) were captured via screen and audio recording, their writing activities and chat protocols

were logged automatically. The students worked on 12.9-inch tablets equipped with external keyboards.

Figure 1

IWM Writing Tool With ChatGPT Integration (Version 4.0)

The primary data source in LiSKIDS 01 consists of think-aloud protocols that capture participants' verbalizations of working memory contents while performing a primary task (Stark, 2010, pp. 58–59). Think-aloud data are often collected for research on social practices (e.g., Noushad et al., 2024), but have played only a minor role in research on writing processes since the work of Hayes and Flower (1980a, 1980b). Criticism of this method includes the potential for reactivity—that is, the effect verbalizing thoughts may have on the processing of the primary task (see Bräuer & Kubik, 2024, p. 301, for a discussion)—as well as the risk of cognitive overload (Kellogg, 2001; Russo et al., 1989). However, researchers disagree on whether the phenomenon of reactivity actually exists. A meta-analysis of 94 studies found that concurrent thinking aloud does not affect the underlying thought process itself (Fox et al., 2011), thereby confirming initial assumptions by Ericsson and Simon (1984) about Level-1 (inner speech) and Level-2 verbalizations (putting nonverbal thoughts into words). LiSKIDS-01 addressed this uncertain situation with the following strategies:

- (a) Preceding data collection, participants received individualized training to familiarize themselves with both the task of thinking aloud and the writing tool. This included viewing a custom-made video demonstrating thinking aloud on a related task and practicing thinking aloud on a similar writing task. Individual training was essential to prevent imitation of peer

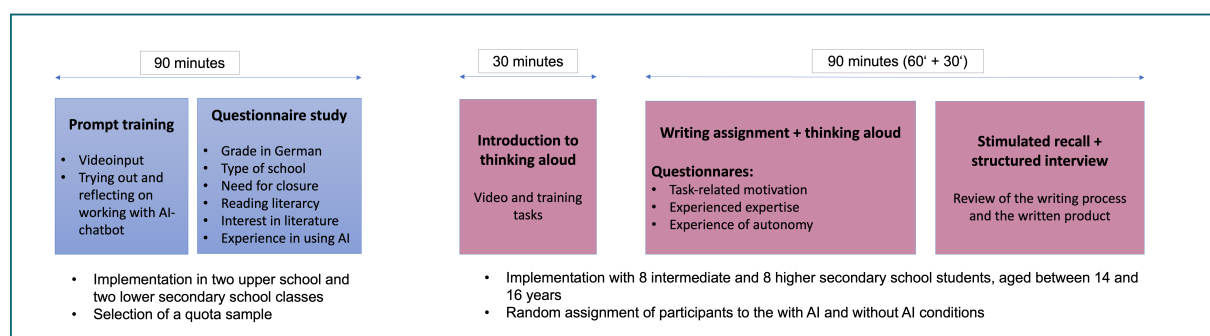
strategies and to enhance participants' confidence in verbalizing their thoughts (Aguado et al., 2013).

- (b) To avoid any additional strain on their cognitive resources, the instructions contained no additional demands (Fox et al., 2011, p. 319).
- (c) In addition to concurrent thinking aloud, subsequent verbal thinking in combination with interrupting the primary writing task was also used when this supported the test subjects, and stimulated recall data (Messmer, 2015) were also collected.
- (d) Although meta-analysis revealed no reactivity for Level-1 and Level-2 verbalizations, it did indicate that metareflective writing strategies and explanations addressed to others could significantly alter results (Fox et al., 2011, p. 333). Therefore, introspection, prompts, or questions that incite test subjects to explain their actions were avoided (Dannecker, 2018; Fox et al., 2011).
- (e) It cannot be ruled out that cognitive and emotional processes can be attributed to the study situation. This aspect of the data must be considered when categorizing and interpreting the results.

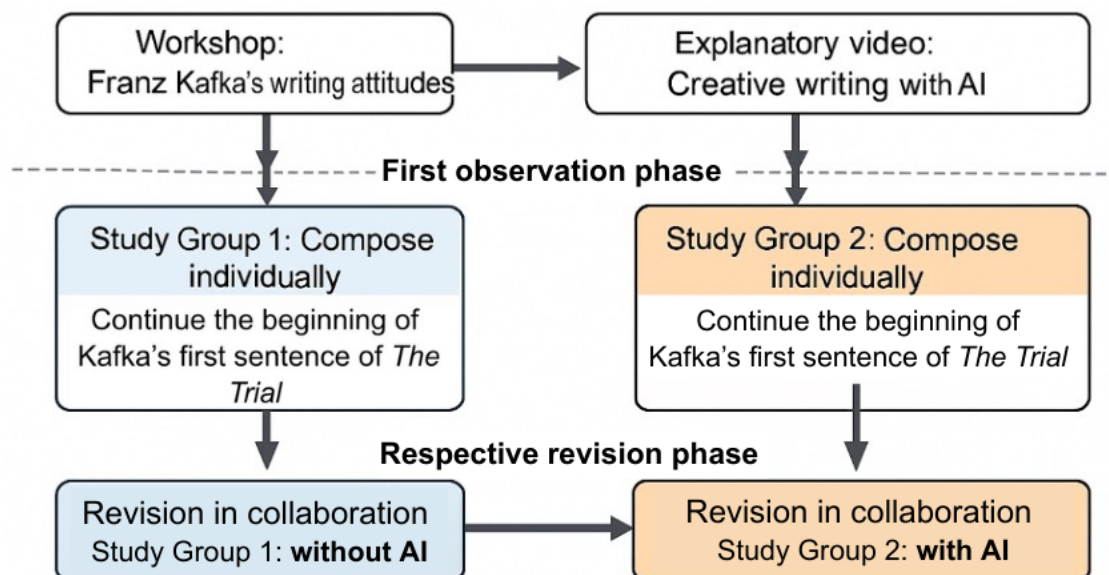
Working with think-aloud data requires further discussion, particularly regarding the epistemological stance of the researchers. Even within the field of cognitive research, it is necessary to consider, for example, whether researchers adhere to the paradigm of traditional information-processing models or to more recent approaches such as the concept of embodied cognition that place greater emphasis on “emotional, social, and institutional dimensions of human information processing” (Schnell, 2013, p. 96, translation D.M.).

Figure 2

Study Design of LiSKIDS 01



In WR-AI-TING, the writing groups used tablets either with or without access to the AI tool. Students in the AI group were free to decide how and when to use the AI tool. Students did not wear headsets because this might have created a feeling of distance and thus distorted results. Several pretests had confirmed that the audio quality was good enough to enable transcription of the revision discussions. The aim was to interfere as little as possible in the students' conversations and to ensure that they could hear each other. Through exchanging their own text drafts, the reflective writing process became authentically visible.

Figure 3*Study Design of WR-AI-TING*

6 Data analysis and interpretation

The LiSKIDS 01 data are evaluated using the paradigm of constructive and reflective grounded theory (Breuer et al., 2019; Charmaz, 2025). The general aim of grounded theory is to establish new theories in a thematic field (Glaser & Strauss, 1967). Therefore, the goal is to expand existing knowledge through hermeneutic methods. This presents researchers with the challenge of deciding how to engage in inductive interpretation processing of the collected data while also applying existing theoretical concepts in a specific research process (Schatzman, 1991). Such research processes should be designed strategically and transparently. Moreover, own perspectives can be expanded through working in interpretative groups. In LiSKIDS 01, the conditions of data collection and evaluation are taken into account, because they constitute a social constellation: (a) First, the situation of data collection reflects a specific social situation constituted by the writing task, the collaboration of an individual with an AI, and the collection situation. (b) Second, the interpretative standpoint of the researchers has to be considered. This does not necessarily have to be understood as a deficit, but can also bear potential as long as the social situatedness of the data and of their interpretation are always taken into account (Bräuer & Kubick, 2024, p. 302). This refers to the important epistemological assumption of LiSKIDS that “social reality is multiple, processual, and constructed” (Charmaz, 2025, p. 12). The aim of the interpretative process is therefore to understand the role and nature of the evaluations that determine the choice of content and formal text design options in early stages of the writing process under different social and communicative conditions. Are these, for example, emotional reactions, intuitions, aesthetic experiences, or specific text norms? Further data collection will be needed to explore literary writing processes (with AI) as a cultural practice (Abraham et al., 2023).

For WR-AI-TING, chat interactions between AI and peers are evaluated by using the documentary method to reconstruct the learners’ conjunctive orientational frameworks (Bohnsack et al., 2013). As

a qualitative method for reconstructing implicit knowledge and collective frames of reference (Bohnsack et al., 2013, p. 12), this approach captures not only the explicit content of the conversations but also the underlying orientations. The underlying assumption is that the orientational frameworks of a group can be reconstructed by analyzing the formal structure of a conversation. It is assumed that this can be documented in the form and manner of an interaction (Asbrand & Martens, 2018, p. 55). Analyses based on the documentary method distinguish two types of knowledge: communicative and conjunctive. The first step is to identify communicative knowledge. This is available to the speakers and is made explicit. Conjunctive knowledge lies one level below this communicative knowledge, and has to be inferred from it (Asbrand & Martens, 2018, p. 50). It is based on conjunctive spaces of experience—that is, people’s shared a priori spaces of experience (Przyborski, 2004, p. 23). The purpose of interpreting data with the documentary method is to reveal the conjunctive understanding of writing processes and especially of revisions—something that is not explicitly apparent from the conversation. The aim is to evaluate the writers’ implicit, collectively shared knowledge, structures, and agencies. Hence, the focus of the data analysis is on communicative and conjunctive knowledge (and the gaps between the two) about creativity, peer cooperation, and interaction with AI as well as metareflections on writing (in terms of ownership, authorship, text production, literariness, etc.).

7 Conclusion

Reconstructive approaches to teaching and learning research, as exemplified by LiSKIDS and WR-AI-TING, provide significant methodological advantages for writing-process research. They enable the comprehensive and nuanced examination of writing processes as situated, social, and meaning-laden events within authentic classroom or learning contexts. Intermediate stages of writing processes, detours, and breaks can be captured accurately and almost authentically (Pelchat, 2022).

The primary objective of LiSKIDS 01 is to formulate theoretical assumptions about the unique features of literary writing processes (with and without AI collaboration) extending beyond existing process models. This encompasses aspects such as idea generation, decision making related to narrative text worlds and their concrete elaboration (such as choice of narrative situation, characters, plot, setting, and time), as well as the negotiation of creative options. The methodological design facilitates the collection of data that at least partially capture the writing process. Alongside think-aloud protocols, this includes drafts and chat transcripts, all of which must be systematically correlated during analysis. Both explicit and implicit literary evaluations can thus be reconstructed by comparing working sequences, thereby enabling the inductive development of theory grounded in the data.

Microprocesses of peer interaction in collaborative writing activities, as investigated in the WR-AI-TING study, can be analyzed by looking at the meaning structures of the actors. The documentary method also takes into account how materialities (digital writing surfaces) and medialities (with or without AI) change and structure writing processes. Through case-specific reconstructions and the comparison of several cases, both individual characteristics of writing processes and typical patterns and structures in collaborative writing settings can be identified. Video-based observations of screen activity and verbal discourse enable researchers to analyze revised text passages as well as the conversation sequences, gestures, and referential practices through which students negotiate meaning in texts. In this way reconstructive classroom research provides systematic access to precisely this link between interaction and text development.

Table 1*Overview*

	LiSKIDS	WR-AI-TING
Test subjects	Aged 14–16 years Intermediate and upper secondary school students	Aged 17–19 years Upper secondary school students
Number of test subjects	16	118
Survey situation	Browser-based online writing tool with ChatGPT integration (Version 4.0)	
	Concurrent and retrospective thinking aloud	Writing discourse
Data analyses	Grounded theory	Documentary method
Writing environment	Individual	Collaborative
	Early writing stage	Revision
conditions	Writing with AI and writing without AI	

References

- Abraham, U., Bräuer, C., & von Brand, T. (Eds.). (2023). *Literarisches Schreiben als kulturelle Praxis: Deutschdidaktische, schreibpädagogische und medienkulturelle Impulse für den Unterricht*. Klett Kallmeyer.
- Aguado, K., Heine, L., & Schramm, K. (Eds.). (2013). *Introspektive Verfahren und qualitative Inhaltsanalyse in der Fremdsprachenforschung*. Peter Lang. <https://doi.org/10.3726/978-3-653-02548-4>
- Asbrand, B., & Martens, M. (2018). *Dokumentarische Unterrichtsforschung*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-10892-2>
- Bohnsack, R., Nentwig-Gesemann, I., & Nohl, A.-M. (2013). *Die dokumentarische Methode und ihre Forschungspraxis: Grundlagen qualitativer Sozialforschung* (3rd ed.). VS Verlag für Sozialwissenschaften.
- Bräuer, C., & Kubik, S. (2024). Lautes Denken als Verbalisationsverfahren literarischer Rezeption in soziokultureller Perspektive – neue Grenzen, neue Chancen. In M.-O. Carl, M. Jörgens, & T. Schulze (Eds.), *Literarische Texte lesen – Texte literarisch lesen* (pp. 281–310). Springer. https://doi.org/10.1007/978-3-662-67816-9_17
- Breuer, F., Muckel, P., & Dieris, B. (2019). *Reflexive grounded theory: Eine Einführung für die Forschungspraxis* (4th ed.). Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-22219-2>
- Charmaz, K. (2025). *Constructing grounded theory*. Sage.
- Dannecker, W. (2018). Lautes Denken. In J. M. Boelmann (Ed.), *Empirische Forschung in der Deutschdidaktik* (Vol. 2, pp. 131–137). Pädagogische Hochschule Ludwigsburg.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. MIT Press.

- Ferdman, A. (2025). AI deskilling is a structural problem. *AI & Society*.
<https://doi.org/10.1007/s00146-025-02686-z>
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, 137(2), 316–344. <https://doi.org/10.1037/a0021663>
- 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), 1–18. <https://doi.org/10.18716/OJS/MIDU/2024.1.3>
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Aldine.
- Hau, M. F. (2025). Writing with machines? Reconceptualizing student work in the age of AI. *Frontiers in Communication*, 10. <https://doi.org/10.3389/fcomm.2025.1598988>
- Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, 29(3), 369–388. <https://doi.org/10.1177/0741088312451260>
- Hayes, J. R., & Flower, L. S. (1980a). Identifying the organization of writing processes. In L. W. Gregg (Ed.), *Cognitive processes in writing* (pp. 3–30). Lawrence Erlbaum Associates.
- Hayes, J. R., & Flower, L. S. (1980b). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg (Ed.), *Cognitive processes in writing* (pp. 31–50). Lawrence Erlbaum Associates.
- Kellogg, R. T. (2001). Competition for working memory among writing processes. *The American Journal of Psychology*, 114(2), 175–191. <https://doi.org/10.2307/1423513>
- Messmer, R. (2015). Stimulated recall als fokussierter Zugang zu Handlungs- und Denkprozessen von Lehrpersonen. *Forum Qualitative Sozialforschung*, 16(1).
- Noushad, B., Van Gerven, P. W. M., & De Bruin, A. B. H. (2024). Twelve tips for applying the think-aloud method to capture cognitive processes. *Medical Teacher*, 46(7), 892–897. <https://doi.org/10.1080/0142159X.2023.2289847>
- Pelchat, L. (2022). *Kollaboratives Schreiben: Eine rekonstruktive Studie zu Problemlösepraktiken im Französischunterricht*. Narr Francke Attempto.
- Philipp, M. (2015). *Schreibkompetenz: Komponenten, Sozialisation und Förderung*. A. Francke Verlag.
- Przyborski, A. (2004). *Gesprächsanalyse und dokumentarische Methode: Qualitative Auswertung von Gesprächen, Gruppendiskussionen und anderen Diskursen*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-90347-7>
- Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, 17(6), 759–769. <https://doi.org/10.3758/BF03202637>
- Schatzman, L. (1991). Dimensional analysis: Notes on an alternative approach to the grounding of theory in qualitative research. In D. R. Maines (Ed.), *Social organization and social process* (pp. 303–314). Aldine de Gruyter.
- Schnell, A. K. (2013). Lautes Denken im Licht der embodied cognition: Neue Grenzen, neue Möglichkeiten in der Schreibprozessforschung? In K. Aguado, L. Heine, & K. Schramm (Eds.), *Introspektive Verfahren und qualitative Inhaltsanalyse in der Fremdsprachenforschung* (pp. 92–115). Peter Lang.
- Siddiqui, M. N., Feliciano, V., Pea, R., & Subramonyam, H. (2025). AI in the writing process: How purposeful AI support fosters student writing. In A. I. Cristea, E. Walker, Y. Lu, O. C. Santos, & S.

- Isotani (Eds.), *Artificial intelligence in education* (Lecture Notes in Computer Science, Vol. 15880). Springer. https://doi.org/10.1007/978-3-031-98459-4_14
- Singh, N., Bernal, G., Savchenko, D., & Glassman, E. L. (2023). Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence. *ACM Transactions on Computer–Human Interaction*, 30(5), 1–57. <https://doi.org/10.1145/3511599>
- Stark, T. (2010). Lautes Denken in der Leseprozessforschung: Kritischer Bericht über eine Erhebungsmethode. *Didaktik Deutsch*, 29, 58–83.
- Steinhoff, T. (2025). Künstliche Intelligenz als Ghostwriter, writing tutor und writing partner: Zur Modellierung und Förderung von Schreibkompetenzen im Zeichen der Automatisierung und Hybridisierung der Kommunikation am Beispiel des Schreibens mit ChatGPT in einer 8. Klasse. In C. Albrecht, J. Brüggemann, T. Kretschmann, & C. Meier (Eds.), *Personale und funktionale Bildung im Deutschunterricht: Theoretische, empirische und praxisbezogene Perspektiven* (pp. 219–244). Metzler.
- Sternberg, R. J. (Ed.). (2003). *Wisdom, intelligence, and creativity synthesized*. Cambridge University Press.
- Sternberg, R. J., & Lubart, T. (1995). An investment approach to creativity: Theory and data. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 262–293). MIT Press. <https://doi.org/10.7551/mitpress/2205.003.0017>
- Torrance, E. P. (1974). *Torrance tests of creative thinking: Verbal tests, forms A and B; figural tests, forms A and B; norms-technical manual*. Personnel Press.

Authors' addresses:

Daniela Matz, University of Tübingen, Deutsches Seminar, Wilhelmstr. 50, 72074 Tübingen, Germany
daniela.matz@uni-tuebingen.de

Carolin Führer, University of Tübingen, Deutsches Seminar, Wilhelmstr. 50, 72074 Tübingen, Germany
carolin.fuehrer@uni-tuebingen.de