e-ISSN 2280-6792

EL.LE
Vol. 15 — Num. 1 — Marzo 2026

Designing an AI-Based Chatbot to Support Autonomous Writing in German as a Foreign Language

Salvatore Varriale
University of Genoa, Italy

Abstract The growing availability of Generative Artificial Intelligence (GenAI) raises new questions about its role in supporting autonomous writing and text revision in second and foreign language learning. This paper presents a design-oriented, quasi-experimental study on a personalised chatbot developed to support autonomous writing in German as a foreign language (DaF). Conducted over one school year with secondary school students at CEFR levels A2-B1, the study compares an experimental group receiving dynamic scaffolding and immediate during-writing feedback with a control group relying on traditional, non-dialogic tools and delayed teacher feedback. Data will include analyses of text quality in supported writing tasks and pre- and post-test tasks without tools, as well as questionnaires and interviews on motivation, self-efficacy and emotions. The study shows how GenAI can function as a process-oriented scaffolding tool that fosters learner autonomy without replacing agency.

Keywords Large Language Models. Generative AI. AI-assisted writing. AI-generated feedback. German as a Foreign Language

Edizioni Ca'Foscari

Peer review

Submitted 2026-02-06
Accepted 2026-03-27
Published 2026-04-27

DOI 10.30687/ELLE/2280-6792/2026/01/005

1 Introduction

The rapid diffusion of Large Language Models (LLMs) has intensified interest in how generative AI can support second language (L2) writing processes. Recent studies indicate that AI-based tools may enhance text revision, learner motivation, and engagement, particularly when they provide timely and task-relevant feedback (Hattie, Timperley 2007; Meyer et al. 2024; Steiss et al. 2024). At the same time, research consistently shows that the pedagogical value of educational technologies depends primarily on their pedagogical design, as technology should not be considered an end in itself (Mishra, Koehler 2006; Celik 2023). In school contexts, writing instruction is often constrained by what has been described as the *bandwidth problem* (Wiley 2006): providing detailed and frequent feedback on student writing requires substantial time and effort, particularly when teachers are responsible for multiple students across several classes. As a result, sustained writing practice and formative feedback are frequently limited (Applebee, Langer 2011; Kiuhara et al. 2009; Graham 2019). These constraints are further compounded outside the classroom, where many students complete writing tasks independently at home without access to individual tutoring or additional instructional support, often due to economic or contextual factors. This paper reports a design-oriented study of a personalised chatbot supporting writing in German as a foreign language (DaF) in secondary education. The chatbot offers dynamic scaffolding and immediate, criteria-referenced feedback while preserving learners' agency. The study examines whether this form of AI-mediated support leads to greater improvements in text revision and writing performance than non-dialogic tools and delayed teacher feedback, and explores students' perceptions in terms of motivation, writing anxiety, and self-efficacy.

2 Theoretical Framework

2.1 Generative AI and Chatbots in Language Education

As LLMs become increasingly powerful and widespread, growing attention has been devoted to their potential role in language education. Recent research highlights the capacity of generative AI to foster learner autonomy, motivation, and language development through engaging and personalised learning experiences that allow learners to progress at their own pace (Allehyani, Algamdi 2023; Annamalai et al. 2023). Within this landscape, chatbots have emerged as particularly promising tools, as they can simulate a wide range of interactions while providing immediate feedback in a

non-judgemental environment, potentially reducing learners' fear of error and evaluation (Barrot 2023; Chiu et al. 2023; Candrasari et al. 2024). Beyond spoken interaction, generative AI tools have also been shown to support vocabulary acquisition, reading comprehension (Kohnke et al. 2023), and written production in L2 contexts (Dos Santos et al. 2023).

2.2 AI-Supported Writing as a Process-oriented Activity

In the domain of writing, AI systems are increasingly conceptualised as collaborative companions that can support different stages of the writing process (Oloff et al. 2022). A systematic review by Wang and Dang (2024) identifies three recurrent forms of AI-assisted writing support: process scaffolding, data analysis, and resource searching. Among these, process-oriented scaffolding emerges as a central pedagogical strategy, defined as guided support that helps learners initiate, connect, and progressively develop ideas through gradual approximation towards independent text production. Such scaffolding encourages learners to conceptualise writing as a process rather than a product and has been shown to increase confidence, especially among L2/FL writers who experience uncertainty during text production. At the same time, a growing body of research underscores that the pedagogical value of AI-assisted writing depends crucially on how AI involvement is designed and perceived (Reza et al. 2025). Learners' sense of ownership and accomplishment is closely tied to their perceived contribution and agency in the writing process (Wang, Wang 2025). Studies show that when AI is framed as an enhancement tool rather than a replacement, learners report stronger feelings of ownership and accomplishment (Kobiella et al. 2023), whereas diminished perceptions of contribution may undermine authorial identity (Rezwana, Maher 2023). These findings highlight the importance of interaction designs that clearly delineate human and AI roles and preserve learner agency.

2.3 Feedback in L2 Writing

Within AI-assisted writing, feedback represents a particularly powerful – yet challenging – pedagogical mechanism. Research on feedback consistently demonstrates that learning outcomes are influenced less by the quantity of feedback than by its timing, focus, and quality (Kluger, DeNisi 1996; Hattie, Timperley 2007). While process-oriented writing instruction with frequent, individualised feedback is empirically well supported (Graham 2018), it remains difficult to sustain in everyday school practice due to time constraints

and workload (Warschauer, Ware 2006; Applebee, Langer 2011). Against this backdrop, AI-mediated feedback has been proposed as a potential means of complementing traditional feedback practices.

In this respect, it is useful to clarify how feedback may differ not only in timing or source, but also in terms of the action taken on the learner's text and the context in which it is provided. Following Li's classification (2025), feedback can be distinguished according to how it addresses the targeted linguistic or textual issue. Action-oriented feedback may involve direct correction, where an erroneous form is explicitly replaced; metalinguistic feedback, which provides comments or clues about the nature of the problem and possible ways to improve it; or indirect feedback, which merely signals the presence of an issue without supplying the correct form.

Other categories include reformulation, whereby a sentence or a larger textual unit is rewritten without altering its meaning, and modelling, which consists in providing an exemplary text on the same task or prompt without intervening directly in the learner's production. Furthermore, Li distinguishes between integrated feedback, which is embedded within the text (e.g., track changes), and detached feedback, which appears as a separate list of comments or a summary at the end of the text. These categories are particularly relevant in the context of AI-generated feedback, as generative tools are capable of producing multiple feedback types – often within the same interaction – ranging from surface-level corrections to more global reformulations or model texts.

2.4 Empirical Evidence on Ai-generated Feedback and Writing Outcomes

Most empirical studies investigating the use of ChatGPT-generated feedback in L2 writing have primarily focused on its impact on writing performance. In a randomised controlled study with upper secondary EFL students, Meyer et al. (2024) found that learners who revised their texts using GPT-generated feedback significantly outperformed peers revising without feedback, showing gains in both revision quality and task motivation. Comparable findings have been reported in higher education contexts (Lo et al. 2025). Similar effects were observed by Alanazi et al. (2025), whose mixed-methods study showed that secondary school students receiving ChatGPT-based automated feedback achieved greater improvements in writing performance than those following traditional instruction, while also reporting generally positive perceptions of the feedback, alongside concerns related to ethical and qualitative aspects. In this regard, further evidence is provided by Koltovskaia et al. (2024), who indicate that L2 students accepted or incorporated approximately 60% of

the revisions suggested by the AI. This selective uptake suggests an evaluative engagement with feedback rather than uncritical adoption. Notably, none of the participants sought supplementary feedback from external sources, such as dictionaries or peers, to further refine their texts.

In contrast to human teachers, who typically adopt a balanced combination of direct and indirect feedback, tools like ChatGPT tend to rely predominantly on metalinguistic feedback and full reformulation. This preference for reformulation can sometimes result in redundant feedback, as the AI may rewrite text that is already grammatically correct merely to alter style. Consequently, scholars argue that, to approximate the pedagogical value of human feedback and preserve learner agency, AI feedback must be carefully designed – for instance, by prompting the system to prioritise indirect or metalinguistic cues over direct correction or reformulation, thereby encouraging learners to actively engage in the revision process rather than passively accepting automated changes.

These differences in feedback practices are also reflected in learners' evaluations and uptake of feedback. When both teacher and GenAI feedback are provided, students tend to evaluate teacher feedback more positively overall and show higher success rates in integrating it into their revisions. Specifically, students engage to a greater extent with GenAI feedback related to text organisation, whereas teacher feedback is favoured and more successfully integrated when it targets content and language-related aspects of writing (Zou et al. 2025). Beyond this distinction, evidence proves that, while AI-generated feedback appears to be highly effective in consistently referencing explicit assessment criteria (Steiss et al. 2024), human feedback remains superior in terms of accuracy, contextual sensitivity, and supportive tone. Taken together, these findings point to important limitations of AI feedback, especially when it is not pedagogically designed or adequately contextualised, as the absence of instructional grounding may reduce its reliability and affective appropriateness. In addition, several studies reveal that GenAI feedback is often perceived as insufficiently individualised, as it may not align with learners' linguistic background, curricular goals, or local assessment practices (Guo, Wang 2024). It may include not only inaccurate but also unnecessary suggestions, fail to recognise irrelevant content, or prioritise surface-level corrections over more substantive issues (Koltovskaia et al. 2024; Lin, Crosthwaite 2024).

Problems of accessibility and usability have also been reported, including confusing presentation formats and difficulty locating feedback within the text (Koltovskaia et al. 2024; Guo, Wang 2024).

2.5 Quality Principles and Accessibility in Ai Feedback Design

The findings discussed above suggest that the value of AI feedback lies not merely in its immediacy or scalability, but in how it is designed and embedded within instruction. For feedback – human or AI-generated – to be genuinely formative, it should adhere to widely shared quality principles (Steiss et al. 2024). Specifically:

- **criteria-referenced feedback**: feedback should explicitly refer to assessment criteria, rubrics, or relevant descriptors (e.g. the CEFR), in order to make task objectives transparent to learners (Black, William 2009; Graham 2018);
- **accuracy**: comments should relate exclusively to what the student has actually written, avoiding additions, inferences, or interpretations that are not supported by the text (Bai, Hu 2017);
- **prioritisation of essential features**: feedback should focus on truly relevant and immediately improvable aspects, rather than overwhelming learners with excessively long or detailed lists of comments (Grimes, Warschauer 2010; Moore, Macharthur 2016);
- **clarity of directions for improvement**: research indicates that effective feedback is clear and uses precise language to provoke actionable writing strategies (Beach, Friedrich 2006; Roscoe et al. 2013);
- **supportive tone:** feedback should maintain an encouraging and non-directive tone, offering suggestions and alternatives while stimulating reflection without embarrassing and/or ridiculing the learner (American Psychological Association Coalition for Psychology in Schools and Education 2015; Graham et al. 2015): this may trigger elevated cortisol levels, which can negatively impact cognitive processes and hinder learning (Balboni et al. 2014).

Within this framework, Li (2025) points out that accessibility represents a crucial criterion for evaluating feedback in L2 writing. Feedback should be delivered in language that learners can readily understand, as students may struggle to process comments formulated using overly sophisticated vocabulary or complex linguistic structures. Unlike L1 contexts, where native-speaking writers typically have no difficulty understanding the language of feedback, accessibility becomes a central pedagogical concern in L2 writing, as the feedback itself may otherwise constitute an additional linguistic barrier rather than a support for learning. An effective prompt to train the chatbot giving feedback should integrate three core components: input, task, and output. The input specifies both the writing task (e.g., prompt, objectives, constraints) and relevant learner characteristics such as age, proficiency level, and educational

context. The task component defines the role assigned to GenAI and the specific action it is expected to perform (e.g., providing feedback as an L2 teacher). Finally, the output component determines the form and focus of AI-generated feedback, including its scope (global vs. local), degree of explicitness, mode of presentation, language of delivery, intended audience, and alignment with assessment criteria or rubrics.

Together, these components ensure that AI feedback is pedagogically aligned, accessible to L2/FL learners, and coherent with instructional goals.

2.6 Literature Gaps and Research Questions

Despite the growing body of research on AI-supported L2 writing, relatively few studies have examined personalised chatbot systems that integrate dynamic scaffolding and immediate feedback throughout the autonomous writing process, particularly within longitudinal, school-based designs that investigate their impact on writing performance while also taking learners' affective responses into account. Much of the existing research has instead relied on short-term interventions (often lasting three or four weeks) or single-session implementations (Wang, Dang 2024), which, while offering valuable insights, leave key dimensions underexplored – most notably the role of timely feedback within an iterative, dialogic writing process (Crosthwaite, Sun 2025), as well as the potential of AI to support autonomous writing practice outside the classroom over extended periods of time.

For instance, Lo et al. (2025) examine AI-generated feedback in a Hong Kong university context, focusing not only on revision practices but also on improvements in overall text quality as the primary indicator of writing performance. Their quantitative analyses show significant gains in essay quality measured through rubric-based assessments of language use, organisation, and content among students receiving AI feedback. However, the reliance on a single writing task and the limited time allocated for feedback consultation and revision constrain the depth and generalisability of the findings. Crucially, the AI system is employed solely as a feedback-generating tool, rather than as an interactive chatbot functioning as a writing partner during the composing process.

Similarly, Meyer et al. (2024) demonstrate that AI-generated feedback can enhance revision quality and task motivation in school-based settings; however, their intervention is limited to a single revision cycle, making it difficult to assess the sustainability of these effects over time. As in Lo et al. (2025), the AI is not designed as a dialogic support system but provides post-hoc feedback on completed

drafts. Finally, research on AI-supported writing has overwhelmingly focused on English as a second or foreign language (Crosthwaite, Sun 2025), with German as a foreign language (DaF) receiving comparatively little attention. On these bases, the present study addresses the following research questions:

- **RQ1:** To what extent do chatbot-supported writing activities influence the quality of written texts in German as a foreign language compared to traditional writing support?
- **RQ2:** How do students perceive AI-supported writing in terms of task motivation, self-efficacy, and writing-related anxiety when compared to traditional forms of writing support and feedback?

3 Chatbot Design and Behaviour

The chatbot, available on both the ChatGPT and Gemini platforms, guides students through the three phases of Zimmerman's self-regulated learning model (1998). Self-regulated learning (SRL) involves the coordinated regulation of cognitive, metacognitive, motivational–emotional, and behavioural dimensions of learning (Persico 2016).

From a cognitive and metacognitive perspective, SRL entails the deliberate use of strategies to process information, solve problems, and reflect on one's own thinking. At the motivational and emotional level, it involves sustaining engagement, managing emotions such as anxiety, and developing self-efficacy beliefs. Behaviourally, SRL includes persistence, adaptive help-seeking, and the active management of the learning environment. Building on this multidimensional view, Zimmerman's model conceptualises SRL as a cyclical process comprising three interrelated phases: forethought, performance, and self-reflection. In the forethought phase, learners analyse task demands, set goals, and plan strategies. During the performance phase, strategies are enacted and regulated through self-monitoring, effort management, and strategic adjustment. In the self-reflection phase, learners evaluate their performance and reflect on the effectiveness of their strategies, informing subsequent learning cycles.

Recent research suggests that LLMs can support SRL by aligning with these regulatory processes. Drawing on Xia et al. (2025), GenAI provides pedagogical affordances that promote learner autonomy by supporting planning, monitoring, and reflection across phases. In the forethought phase, AI tools can assist learners in clarifying objectives and planning their work proactively (Hew et al. 2023). During the performance phase, timely feedback and information support can facilitate self-monitoring and strategic adjustment while reducing

cognitive load (Chen 2024). In the self-reflection phase, analysis of learning behaviours can foster metacognitive awareness and support self-evaluative judgments, preparing learners for subsequent learning cycles (Sun et al. 2023; Takagi et al. 2023).

In line with this framework, the chatbot operationalises SRL support across the three phases of Zimmerman's model, while allowing flexible movement between phases in line with Flower and Hayes' process-oriented model of writing (1981). More specifically, during the planning phase, the chatbot poses guiding questions such as "Why did you choose this topic?" and may suggest simple outlines or conceptual maps to help structure the text before writing.

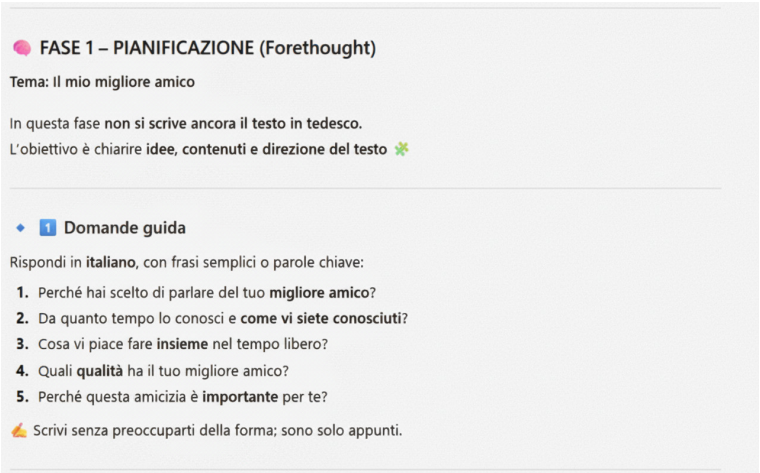

Figure 1 Guiding questions for idea generation and planning of a writing task in German as a Foreign Language (own illustration)

When talking about friendship **[fig. 1]**, and in particular about a best friend, the chatbot prompts students with guiding questions such as how long they have known each other, what they do together in their free time, why this friendship is important, and which qualities the friend has. Importantly, the chatbot explicitly reassures students that they should not worry about linguistic form at this stage, as the focus is on generating ideas and taking notes. In the performance phase, the chatbot first asks students to produce an initial draft of the text. Upon request, the chatbot can also provide lists of topic-related vocabulary or expressions. These lists were developed in collaboration with the students' teachers and are aligned with the Common European Framework of Reference for Languages (CEFR). As shown in the example conversation **[fig. 2]**, before inviting students to write their draft, the chatbot offers a set of writing strategies.

These include encouraging the use of brief and clear sentences, the use of connectives to foster textual cohesion, and the explicit motivation of ideas. Finally, the chatbot prompts students to check whether their sentences are complete and explicitly emphasises that the text they are producing is only a draft. The chatbot then provides in-process feedback in order to foster monitoring and self-control. Students are subsequently invited to revise their text and submit a final version.

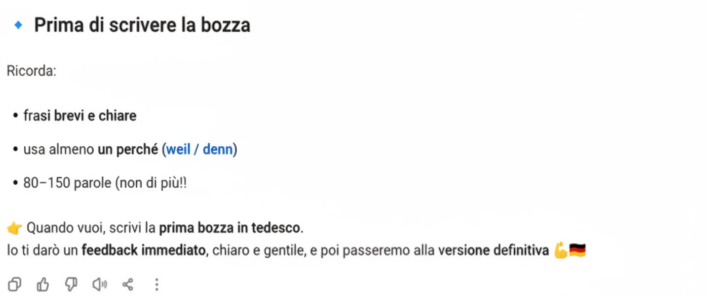

Figure 2 Guidelines for drafting a short text in German as a foreign language (own illustration)

Feedback on the final version of the text serves as a bridge to the reflection phase. In this phase **[fig. 3]**, students are invited to reflect on their writing process through a brief metacognitive diary, responding to questions such as “What difficulties did you encounter?” or “Which strategies did you use to overcome them?”. By explicitly prompting learners to articulate challenges and strategies, this interaction is designed to foster metacognitive awareness and to support the self-reflection phase of self-regulated learning, encouraging learners to consolidate insights gained during writing and revision for future tasks.

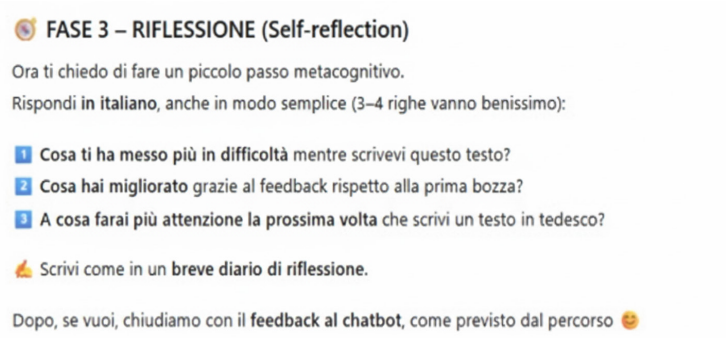

Figure 3 Guiding questions for self-reflection on difficulties, learning progress, and future writing strategies (own illustration)

Behind the chatbot's design and behaviour lies the teacher's pedagogical expertise; as highlighted by Celik (2023), the Intelligent-TPACK framework emphasises teachers' capacity to design purposeful prompts that align generative AI with instructional goals. Since output quality depends not only on the model or its training data but also on prompt effectiveness, the instructional design of the chatbot is strongly based on the principles of prompt engineering, understood as the practice of crafting precise textual inputs to guide AI responses. An effective prompting strategy involves starting with simple instructions and progressively adding detail and contextual information based on the responses obtained. To this end, ambiguity in natural language must be avoided as much as possible, and prompts should be clear and descriptive, with examples provided when necessary. Unlike human communication, where politeness plays a central role, interaction with GenAI requires a strong emphasis on clarity and conciseness (Dornburg, Davin 2024; Giray 2023). It is therefore advisable to use direct action verbs (e.g. *write, classify, summarise, translate*) to formulate precise and unambiguous instructions.

Building on these principles, the chatbot's behaviour is regulated through a set of explicit prompting guidelines embedded directly in the system prompt. At a general level, the chatbot is instructed to maintain a polite, encouraging, and clear interactional style, while avoiding excessive positive reinforcement that could reduce the instructional value of feedback. Its language is systematically adapted to learners aged 16-18 and tailored to the CEFR A2/B1 level, ensuring accessibility without oversimplification. Crucially, the prompting guidelines strictly prohibit the chatbot from performing the task on behalf of the learner: the system is instructed never to generate complete drafts or model texts, nor to engage in direct

rewriting of students' productions. Instead, feedback is designed to operate primarily through indirect and metalinguistic strategies (Li 2025), prompting learners to identify and revise problematic areas autonomously.

In a first step, the chatbot visually highlights all potential errors in bold, with a particular focus on orthographic and morphological features. This initial intervention constitutes a form of indirect feedback, as deviations are signalled throughout the text without being explicitly corrected or reformulated. By increasing the perceptual salience of problematic forms, the chatbot prompts learners to notice errors and engage in self-correction, while maintaining their responsibility and agency in the writing process. In a second step, the extended feedback adopts a metalinguistic focus, selectively addressing only two or three key issues that are considered most important and immediately improvable. The feedback is metalinguistic because it does not merely point out incorrect forms but explicitly draws learners' attention to the underlying linguistic rules involved, for example subject–verb agreement, plural formation, or the systematic capitalisation of nouns in German, a recurrent source of difficulty for Italian learners of DaF. Through brief explanations, reminders, and illustrative examples, the chatbot supports learners in reflecting on these rules and applying them autonomously in revision.

This approach directly addresses a recurrent issue in AI-assisted L2 writing identified in the systematic review by Wang and Dang (2024), namely the tendency of GenAI tools to automatically edit or rewrite students' original texts. Such practices may encourage copy-paste revision and limit opportunities for reflection and active decision-making during revision, with apparent improvements in the final product that reflect short-term, text-bound gains rather than genuine development of writing competence. All of this is carried out in line with widely shared quality principles for formative feedback (Steiss et al. 2024), which, for example, recommend focusing on a limited number of truly relevant and immediately improvable aspects, rather than overwhelming learners with excessively long or detailed lists of comments (Grimes, Warschauer 2010; Moore, MacArthur 2016). The prompt further instructs the chatbot to encourage the use of short, clear, and simple sentences, as well as the use of connectors such as *denn* and *weil* to express reasons, with the explicit motivation that doing so may lead to the awarding of additional points. To ensure task consistency and comparability across sessions, the chatbot monitors compliance with predefined constraints, including a required text length of 80-150 words. Taken together, these prompting guidelines ensure that the chatbot serves as a scaffolding-oriented instructional tool rather than a text generator, thus preserving learner agency.

4 Methodology

The intervention will last for one school year and will be conducted at the Liceo Linguistico Internazionale Grazia Deledda in Genoa. It will involve students with proficiency levels ranging from A2 to B1 according to the Common European Framework of Reference for Languages (CEFR). All writing tasks will be selected from the *Deutsches Sprachdiplom* (DSD) task catalogue and will focus on informative text types, such as writing about hobbies, family, media, and related everyday topics. The instructional intervention is structured into seven sessions distributed over the school year: an initial in-class session (Session 1) to establish a baseline, a sequence of five guided writing sessions carried out at home (Sessions 2-6), and a final in-class session (Session 7) to examine potential transfer effects. Two groups will be compared: an experimental group (n = 16) using a personalised chatbot that provides immediate feedback, and a control group (n = 22) using metacognitive checklists and other traditional, non-dialogic writing support tools, with feedback provided by the teacher at a later stage. Data collection focuses on two main dimensions:

- **Quality of text revision and writing performance (quantitative analysis)**: In Sessions 2 to 6, students in the experimental group will complete the writing tasks individually at home using the chatbot: they will produce an initial draft, receive immediate, feedback after draft submission within an iterative writing process, and submit a revised version. The final text will be submitted by sharing a chatbot-generated link, which students will send to the teacher via email. Writing improvement will be operationalised by calculating the difference (delta) between the initial draft and the revised text using an analytic scoring rubric aligned with CEFR A2–B1 descriptors, assessing text complexity, linguistic accuracy, and textual cohesion and coherence. The control group will likewise complete the writing tasks individually at home, working on a shared document (e.g. Google Drive) whose structure mirrors that of the chatbot-based sessions. Students will submit an initial draft, receive delayed teacher feedback, and then produce a revised version. During the writing process, learners may use dictionaries, grammar overviews, and other traditional support tools. Improvement will be calculated for this group using the same analytic rubric and delta procedure. In both groups, learners will have two days to revise their texts after receiving feedback. In contrast, Sessions 1 and 7 will be conducted in class, where all students will produce written texts without any support tools in order to assess writing quality under conditions of full autonomy and to

examine potential transfer effects beyond the supported writing phases.

- **Students' perceptions (quantitative and qualitative analysis):** Students' perceptions are investigated through Likert-scale questionnaires and semi-structured interviews, exploring how learners experience different types of support and feedback in relation to task motivation, self-efficacy, and writing anxiety.

5 Preliminary Baseline Results

Prior to the first writing task, a baseline questionnaire on writing-related motivation, self-efficacy and anxiety was administered to both groups. Results show high motivation to improve writing in German and high anxiety in evaluative contexts, alongside comparatively low intrinsic interest in writing.

The experimental group reported greater difficulty with independent writing and a higher fear of negative teacher evaluation. Open-ended responses from both groups reveal similar perceived difficulties, mainly related to grammatical accuracy (verb placement, case, word order), lexical choice and text organisation. With regard to feedback practices, participants in the experimental group in particular highlighted the benefits of immediately available feedback for reducing uncertainty during writing. Overall, learners in both groups expressed a positive but critical attitude towards AI, viewing it as a supportive tool for brainstorming, planning and revision rather than as a replacement for autonomous writing.

6 Conclusion

This paper has presented a design-oriented investigation of a personalised chatbot developed to support autonomous writing in German as a foreign language in secondary school contexts. Grounded in models of self-regulated learning and process-oriented writing, the chatbot was designed to provide dynamic scaffolding and immediate, criteria-referenced feedback while deliberately preserving learners' agency and responsibility for text production.

Through a longitudinal comparison between AI-supported writing and traditional, non-dialogic tools with delayed teacher feedback, the study addresses key gaps in existing research, particularly in relation to school-based contexts, adolescent learners, and languages other than English. Specifically, it investigates how repeated exposure to a personalised chatbot acting as a dialogic writing partner may be associated with improvements in text revision and writing performance as well as with changes in task motivation, self-efficacy,

and writing-related anxiety, responding to recent calls for a more nuanced understanding of AI-based educational tools.

The analysis of students' initial texts (T1), intended as a diagnostic reference point, will allow recurrent error patterns and writing difficulties to be mapped through a systematic error analysis. Errors will be grouped into broader categories following the classification proposed by Mlakar Gračner (2018), namely morpho-syntactic errors, lexico-semantic errors, and orthographic and punctuation errors. With regard to textuality, the category of discourse-related errors as defined by Chlosta et al. (2010) will also be adopted. This analysis will inform the progressive calibration of the prompts and feedback requests addressed to the chatbot in response to learners' emerging needs. In addition, a focus group with teachers will support the validation of the tool by offering practitioner-based insights into its pedagogical adequacy and by identifying directions for its further development. Beyond the specific research questions addressed, the study outlines a pedagogically grounded chatbot design that may inform future research and practice in FL writing instruction. While broader implications for instructional innovation and teacher professional development lie beyond the scope of the present analysis, the proposed model highlights the central role of pedagogical design in supporting formative feedback, learner autonomy, and reflective writing practices in AI-assisted language education.

This perspective aligns with recent research by Saeli et al. (2025), who investigated EFL teachers' attitudes toward adopting ChatGPT as a teaching and feedback tool through a survey-based study. Their findings reveal a generally high level of openness toward professional development related to AI use, with the highest-rated item indicating teachers' strong willingness to invest time and effort in learning how to use ChatGPT effectively. Such results highlight the importance of pedagogically informed models that not only introduce AI tools into language classrooms but also support teachers in developing the necessary competences to integrate them responsibly and meaningfully.

Bibliography

Alanazi, S.M. et al. (2025). "Assessing the efficacy of ChatGPT's automated corrective feedback in enhancing students' writing proficiency". *International Journal of Advanced and Applied Sciences*, 12(2), 205-14. https://doi.org/10.21833/ijaas.2025.02.022.

Allehyani, S.H.; Algamdi, M.A. (2023). "Digital competences: Early childhood teachers' beliefs and perceptions of ChatGPT application in teaching English as a second language (ESL)". *International Journal of Learning, Teaching and Educational Research*, 22(11), 343-63.

American Psychological Association, Coalition for Psychology in Schools and Education (2015). "Top 20 principles from psychology for PreK-12 teaching and learning". https://www.apa.org/ed/schools/cpse/top-twenty-principles.pdf.

Annamalai, N. et al. (2023). "Exploring English Language Learning Via Chatbot: A Case Study from a Self-determination Theory Perspective". *Computers and Education: Artificial Intelligence*, 5, 100148, 1-8. https://doi.org/10.1016/j.caeai.2023.100148.

Applebee, A.N.; Langer, J.A. (2011). "EJ Extra: A Snapshot of Writing Instruction in Middle Schools and High Schools". *English Journal*, 100(6), 14-27. https://doi.org/10.58680/ej201116413.

Bai, L.; Hu, G. (2017). "In the face of fallible AWE feedback: How Do Students Respond?" *Educational Psychology*, 37(1), 67-81. https://doi.org/10.1080/01443410.2016.1223275.

Balboni, P.E.; Caon, F.; Daloiso, M. (2014). *Come funzionano il cervello e la mente nell'acquisizione linguistica*. Venezia: Università Ca' Foscari Venezia, 1-12.

Barrot, J.S. (2023). "Using ChatGPT for second language writing: Pitfalls and potentials". *Assessing Writing*, 57, 100745, 2. https://doi.org/10.1016/j.asw.2023.100745.

Beach, R.; Friedrich, T. (2006). "Response to Writing". *Handbook of writing research*. New York: Guilford.

Black, P.; Wiliam, D. (2009). "Developing the Theory of Formative Assessment". *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5.

Candrasari, R.; Rahman, A.; Nurhaliza, S. (2024). "GPT Chat: Useful or not in Supporting Learning in Higher Education". *International Journal of Language and Ubiquitous Learning*, 2(2), 113-25.

Celik, I. (2023). "Towards intelligent-TPACK: An Empirical Study on Teachers' Professional Knowledge to Ethically Integrate Artificial Intelligence (AI)-based tools into education". *Computers in Human Behavior*, 138, 107644. https://doi.org/10.1016/j.chb.2022.107468.

Chen, M.R.A. (2024). "Metacognitive mastery: Transformative learning in EFL through a generative AI chatbot fueled by metalinguistic guidance". *Educational Technology & Society*, 27(3), 407-27.

Chiu, T.K.F.; Moorhouse, B.L.; Chai, C.S.; Ismailov, M. (2024). "Teacher Support and Student Motivation to Learn with Artificial Intelligence (AI) Based Chatbot". *Interactive Learning Environments*, 32(7), 3240-3256. https://doi.org/10.1080/10494820.2023.2172044.

Chlosta, C.; Schäfer, A.; Baur, R.S. (2010). "Fehleranalyse". Ahrenholz, B; Oomen-Welke, I. (Hrsgg), *Deutsch als Zweitsprache.* Bd. 9, *Deutschunterricht in Theorie und Praxis*. Baltmannsweiler.

Crosthwaite, P.; Sun, S. (2025). "Generative AI and L2 written feedback studies: A scoping review". *RELC Journal*, 57(1). https://doi.org/10.1177/00336882251386530.

Dornburg, A.; Davin, K.J. (2024). "ChatGPT in foreign language lesson plan creation: Trends, variability, and historical biases". *ReCALL*, 3-4. https://doi.org/10.1017/S0958344024000272.

Dos Santos, A.E. et al. (2023). "ChatGPT in ELT: Writing affordances and activities". *TESOL Connections*, 1-6.

Flower, L.; Hayes, J.R. (1981). "A Cognitive Process Theory of Writing". *College Composition and Communication*, 32(4), 365-87. https://doi.org/10.2307/356600.

Giray, L. (2023). "Prompt Engineering with Chatgpt: A Guide for Academic Writers". *Annals of Biomedical Engineering*, 51, 2629-33.

Graham, S. (2018). "Instructional feedback in writing". Lipnevich, A.A.; Smith, J.K. (eds), *The Cambridge handbook of instructional feedback*. Cambridge: Cambridge University Press, 145-68.

Graham, S. (2019). "Changing How Writing is Taught". *Review of Research in Education*, 43(1), 277-303. https://doi.org/10.3102/0091732X18821125.

Graham, S.; Hebert, M.; Harris, K.R. (2015). "Formative Assessment and Writing: A Meta-Analysis". *The Elementary School Journal*, 115(4), 523-47. https://doi.org/10.1086/681947

Grimes, D.; Warschauer, M. (2010). "Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation". *Journal of Technology, Learning, and Assessment*, 8(6).

Guo, K.; Wang, D. (2024). "To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing". *Education and Information Technologies*, 29(7), 8435-8463. https://doi.org/10.1007/s10639-023-12146-0.

Hattie, J.; Timperley, H. (2007). "The Power of Feedback". *Review of Educational Research*, 77(1), 81-112. https://doi.org/10.3102/003465430298487.

Hew, K.F.; Huang, W.; Du, J.; et al. (2023). "Using Chatbots to Support Student Goal Setting and Social Presence in Fully Online Activities". *Journal of Computing in Higher Education*, 35, 40-66. https://doi.org/10.1007/s12528-022-09338-x.

Kiuhara, S.A.; Graham, S.; Hawken, L.S. (2009). "Teaching Writing to High School Students: A National Survey". *Journal of Educational Psychology*, 101(1), 136-60.

Kluger, A.N.; DeNisi, A. (1998). "Feedback Interventions: Toward the Understanding of a Double-edged Sword". *Current Directions in Psychological Science*, 7(3), 67-72. https://doi.org/10.1111/1467-8721.ep10772989.

Kobiella, C. et al. (2024). "'If the machine is as good as me, then what use am I' How the use of ChatGPT changes young professionals' perception of productivity and accomplishment". *Proceedings of the CHI Conference on Human Factors in Computing Systems*. New York: ACM. https://doi.org/10.1145/3613904.3641964

Kohnke, L.; Moorhouse, B.L.; Zou, D. (2023). "ChatGPT for language teaching and learning". *RELC Journal*, 54(2), 537-50. https://doi.org/10.1177/00336882231162868.

Koltovskaia, S.; Rahmati, P.; Saeli, H. (2024). "Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement". *Journal of Second Language Writing*, 65, 1-15.

Li, S. (2025). "Generative AI and Second Language Writing". *Digital Studies in Language and Literature*, 2(1), 122-52. https://doi.org/10.1515/dsll-2025-0007.

Lin, S.; Crosthwaite, P. (2024). "The Grass is not Always Greener: Teacher vs. GPT-Assisted Written Corrective Feedback". *System*, 127, 103529.

Lo, N.; Wong, A.; Chan, S. (2025). "The impact of generative AI on essay revisions and student engagement". *Computers and Education Open*, 9, 100249. https://doi.org/10.1016/j.caeo.2025.100249.

Meyer, J.; Jansen, T.; Schiller, R.; Liebenow, L.W.; Steinbach, M.; Horbach, A.; Fleckenstein, J. (2024). "Using LLMs to bring evidence-based feedback into the classroom". *Computers and Education: Artificial Intelligence*, 6, 100199. https://doi.org/10.1016/j.caeai.2023.100199.

Mishra, P.; Koehler, M.J. (2006). "Technological Pedagogical Content Knowledge". *Teachers College Record*, 108(6), 1017-54.

Mlakar Gračner, D. (2018). "Fehler beim schriftlichen Produzieren von Texten im DaF-Studium. Eine empirische Untersuchung am Beispiel slowenischer Germanistikstudierender". *Vestnik za tuje jezike/Journal for Foreign Languages*, 10(1).

Moore, N.S.; MacArthur, C.A. (2016). "Student Use of Automated Essay Evaluation Technology During Revision". *Journal of Writing Research*, 8(1), 149-75.

Oloff, C.; Biermann, N.F.; Ma, F.N.; Yoon, D. (2022). "From Tool to Companion". *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. New York: ACM, 1209-27.

Reza, M. et al. (2024). *Co-Writing with AI, on Human Terms: Aligning Research with User Demands Across the Writing Process.* Toronto: University of Toronto.

Rezwana, J.; Maher, M.L. (2023). "User Perspectives on Ethical Challenges in Human-AI Co-Creativity". *Proceedings of the 15th Conference on Creativity and Cognition*. New York: ACM, 62-74. https://doi.org/10.1145/3591196.3593364.

Roscoe, R.D.; Varner, L.K.; Crossley, S.A.; McNamara, D.S. (2013). "Developing pedagogically-guided algorithms for intelligent writing feedback". *International Journal of Learning Technology*, 8(4), 362-81.

Saeli, H. et al. (2025). "Embrace or resist? EFL teachers on adopting ChatGPT as a teaching and feedback tool". *Language Awareness*, 2-18. https://doi.org/10.1080/09658416.2025.2577191.

Schunk, D.H.; Zimmerman, B.J. (1998). *Self-regulated learning*. New York: Guilford Press.

Steiss, J. et al. (2024). "Comparing the quality of human and ChatGPT feedback". *Learning and Instruction*, 91, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894.

Sun, J.C.-Y.; Tsai, H.-E.; Cheng, W.K.R. (2023). "Effects of integrating an open learner model with AI-enabled visualization". *Computers and Education: Artificial Intelligence*, 4, 100120.

Takagi, M. et al. (2023). "Development and evaluation of a support system for creating learning plans". *Journal of Information Processing*, 31, 440-51.

Wang, H.; Dang, A. (2024). *Enhancing L2 writing with generative AI: A systematic review of pedagogical integration and outcomes*. Tucson: University of Arizona, 12.

Wang, C.; Wang, Z. (2025). "Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model". *Journal of Second Language Writing*, 67, 101187, 2. https://doi.org/10.1016/j.jslw.2025.101187.

Warschauer, M.; Ware, P. (2006). "Automated Writing Evaluation". *Language Teaching Research*, 10(2), 157-80.

Wiley, D.A. (2005). "Learning objects in public and higher education". Spector, J.M.; Ohrazda, C.; Van Schaack, A.; Wiley, A.D. (eds), *Innovations in Instructional Technology: Essays in Honor of M. David Merrill*. Mahwah (NJ); London: Lawrence Erlbaum Associates, 1-9.

Xia, Q. et al. (2025). "A Systematic Literature Review on Designing Self-Regulated Learning Using Generative AI". *Computers & Education*, 240, 105465, 2-9. https://doi.org/10.1016/j.compedu.2025.105465.

Zou, S. et al. (2025). "Investigating Students' Uptake of Teacher- and ChatGPT-Generated Feedback in EFL Writing: A Comparison Study". *Computer Assisted Language Learning*, 1-30. https://doi.org/10.1080/09588221.2024.2447279.