



How to support at-risk writers: Differential effects of formative feedback on argumentative writing and motivation

Katrin Peltzer¹ · Alina Lira Lorca² · Ulrike-Marie Krause² · Steve Graham³ · Ernesto Panadero^{4,5} · Vera Busse¹

Accepted: 12 March 2025 / Published online: 15 April 2025
© The Author(s) 2025

Abstract

Formative feedback fosters writing and student motivation, but differential effects on writers with varying achievement levels are under-researched. It remains unclear to what extent time-efficient methods such as rubric and exemplar feedback support less-proficient writers. Our study addresses this gap by comparing the effects of different formative feedback methods on deep-level features in writing, self-efficacy, enjoyment, and feedback perceptions in two groups of secondary school EFL students: students designated at-risk of writing failure ($n = 101$) and more proficient developing writers ($n = 101$). We conducted a randomized controlled intervention study with four conditions: EG1 (rubric+exemplar feedback), EG2 (in-text comments), EG3 (rubric+exemplar feedback and in-text comments), and CG1 (learning unit without additional feedback). Findings showed that rubric+exemplar feedback had a differential effect, particularly benefitting at-risk writers. In both achievement-level groups, in-text comments (EG2) led to comparable learning progress as in the control group without additional feedback (CG1). Importantly, more feedback (EG3) only had additional benefits for developing writers, while at-risk writers in EG3 made similar progress to at-risk writers in the control condition. At-risk writers had lower self-efficacy and writing enjoyment, with no significant changes over time, while developing writers experienced slight losses in enjoyment. Both at-risk and developing EFL writers had positive perceptions of all feedback types, though the motivational impact of the feedback was rated lower than its quality. In conclusion, although we did not observe motivational changes, rubric+exemplar feedback seems helpful for at-risk writers and is positively received by students.

Keywords Individual differences · EFL · Writing self-efficacy · Writing enjoyment · Rubric · Exemplar

Introduction

Formative feedback, which provides guidance and highlights areas for improvement, is widely recognized as an essential tool for fostering writing development, and its specific benefits for language learners have been amply documented (see the meta-analyses by Biber et al., 2011; Scherer et al., 2024). Yet the effects of feedback on learner subgroups, especially foreign language (FL) learners at risk of writing failure, have received limited attention. Such at-risk writers may be particularly affected by both motivational (see Mills, 2014; Pajares, 2003) and cognitive (see Galbraith, 2009) challenges. Affective-motivational factors influence writers' engagement and perseverance in the writing process (Graham, 2018), which may explain why at-risk writers tend to avoid writing exercises instead of writing more (Woodrow, 2011). Regarding cognitive challenges, research suggests that developing deep-level features, such as creating coherence and meaningful structure in texts, poses a difficulty, especially in genres that require complex organizational skills, like argumentative writing (Siekman et al., 2022).

In many classroom settings, formative feedback is traditionally provided as in-text comments. However, more recently, time-efficient feedback methods, such as rubric-based and exemplar feedback, have gained increased attention, suggesting a significant potential to support learners (Lipnevich et al., 2014; Panadero et al., 2023; To et al., 2022), particularly in large classes. In a previous study, we found that rubric and exemplar feedback effectively helped FL students develop argumentative writing and improved holistic writing quality (Peltzer et al., 2024).

However, using rubric and exemplar feedback to improve texts also requires students to independently engage with the criteria, and little is known whether such feedback addresses the specific needs of at-risk writers. The latter may need more direct instruction and encouragement and might, therefore, perceive in-text comments more favorably. In particular, in-text comments may be better suited to support writing self-efficacy and enjoyment among at-risk writers.

Our randomized controlled feedback intervention examines differential effects on learners with varying achievement levels. It responds to general calls for more studies on feedback addressing deep-level features of writing among language learners (Scherer et al., 2024), considers feedback perceptions (Berry et al., 2025), and addresses the lack of research on the motivational impact of different feedback practices (Camacho et al., 2021).

Theoretical and empirical background

Cognitive challenges of foreign language writing and revision

Language learners often face cognitive challenges affecting writing, such as increased cognitive load (Abu-Rabia, 2003; Johnson, 2020), which can impair

planning (Silva, 1993) and hinder coherent, well-structured text production (Hinkel, 2011; Silva, 1993). There is also empirical evidence that language learners often write in shorter bursts and display more frequent language revisions than students writing in their L1 (Chenoweth & Hayes, 2003).

Reviewing established L1 writing models (such as Bereiter & Scardamalia, 1987; Chenoweth & Hayes, 2003; Hayes, 1996; Hayes & Flower, 1980; Kellogg, 1988), Galbraith (2009) notes that writing in a less familiar language is more effortful as the stages of planning, translating, and revising demand more cognitive resources due to limited linguistic proficiency and less automated writing processes. As a result, language learners often struggle to focus on higher-order concerns like coherence and structure. This difficulty is further compounded in genres requiring high levels of coherence, like argumentative writing. In comparison to other genres, such as narrative writing, language learners at secondary school frequently struggle to link ideas and maintain structural clarity in argumentative texts, often omitting entire sections, such as conclusions (Siekman et al., 2022). At the same time, argumentative writing is often a key component of writing assessments at secondary school (for the US context, see Newell et al., 2011; for a discussion of German curricular requirements for EFL writing, see Peltzer et al., 2022).

Apart from planning, revision is crucial for improving writing, yet difficult to master for many students (Graham & Perin, 2007). Flower and colleagues (1986) emphasize that, in general, effective revision involves aligning a text with the writer's goals, which is inherently complex and demands significant training (Fitzgerald, 1987; Flower et al., 1986). It is thus crucial to support writers throughout the demanding revision process, particularly FL writers, who tend to focus on surface-level revisions due to lower proficiency (Stevenson et al., 2006).

Feedback to support less-proficient English as a foreign language (EFL) writers

Feedback can support revision by guiding students' attention to deep-level features of texts (Hyland, 2003) and exerts considerable influence on writing motivation (Busse, 2013; Camacho et al., 2021). However, the effects of feedback on writing and motivational outcomes may depend on feedback characteristics (e.g., form, content, timing) and students' language proficiency (Biber et al., 2011; Scherer et al., 2024; see also Busse et al., 2024). An early meta-analysis (Biber et al., 2011) showed that language learners ($d = 1.53$) benefited more from feedback than learners writing in their L1 ($d = 1.20$), with low-proficiency language learners profiting more ($d = 1.35$) than their high-proficiency peers ($d = 1.03$). That meta-analysis further suggested that combining feedback on both content (e.g., text coherence) and form (e.g., spelling and grammar) is more effective for learning and more motivating than focusing solely on formal errors (Biber et al., 2011). In a more recent meta-analysis, Scherer and colleagues (2024) analyzed differential effects of feedback on L1, L2, and FL learners, distinguishing between surface-level (e.g., spelling and grammar) and deep-level features (e.g., structure, coherence). Deep-level feedback alone positively impacted deep-level features in FL learners ($g = 0.37$), while a combination of surface- and deep-level feedback effectively enhanced both surface- ($g = 0.54$) and

deep-level outcomes ($g=0.82$). Conversely, surface-level feedback alone negatively affected deep-level outcomes ($g=-0.23$). Notably, most studies included in this meta-analysis were conducted in tertiary education, and none investigated the effects of deep-level instructor feedback in secondary school.

Regarding feedback characteristics, effective feedback should be integrated into the learning process (Panadero & Lipnevich, 2022; Shute, 2008) and provide clear information on current progress (*feed back*), learning objectives (*feed up*), and next steps (*feed forward*) (Hattie & Timperley, 2007). Novices or struggling writers especially benefit from clear guidance (Shute, 2008). However, feedback practices effective for L1 writers may not equally suit language learners; for example, a study by Müller and colleagues (2023) revealed that teacher feedback including feed forward was a significant predictor of text quality for L1 writers but not for multilingual writers with migration backgrounds.

Timely feedback is vital, especially for weaker writers (Shute, 2008), and focusing on deep-level features early promotes better revision (McGarrell & Verbeem, 2007). However, providing timely feedback that adheres to quality standards presupposes that feedback providers have ample resources, which is often not the case. In particular, traditional written comments in essay margins are time-consuming and may overwhelm students with excessive revision instructions (see also Busse, 2013; F. Hyland & Hyland, 2001). Therefore, developing alternative feedback methods that support self-directed revision and enhance feedback literacy is crucial.

A more time-efficient feedback method involves using rubrics and exemplars (Lipnevich et al., 2023). Rubrics outline task criteria and fulfillment levels (Brookhart, 2018), while exemplars, or model texts, demonstrate optimal performance, helping students visualize a good essay. Both clarify outcome expectations, making them more understandable (Lipnevich et al., 2023), and potentially encourage independent exploration of quality criteria (Carless & Boud, 2018) and self-evaluation against established standards (Lipnevich et al., 2023; Nicol, 2021). Clear rubrics are thus a valuable tool for reducing cognitive load (Lipnevich et al., 2023). However, in German secondary English as a foreign language (EFL) classrooms, where rubric and exemplar feedback is infrequently used (Siekmann & Busse, 2022), students may be unfamiliar with this type of feedback. Therefore, further investigation is needed to determine if this feedback effectively supports EFL students with low language proficiency, especially those at risk of writing failure.

Variables related to writing influenced by feedback

Deep-level features in argumentative writing

Defining good texts as “coherently organized essays containing well-developed and pertinent ideas, supporting examples, and appropriate detail” (Graham & Perin, 2007, p. 14, referencing Needels & Knapp, 1994) highlights the importance of deep-level features. These features encompass a text’s underlying components such as structure, coherence, and audience orientation (Parr & Timperley, 2010).

In argumentative writing, these features are essential for presenting a persuasive argument. A well-structured argumentative essay typically includes an introduction with a thesis statement, a main body with supporting arguments, and a conclusion (Hyland, 1990). Toulmin's argument model (2003) additionally identifies six argumentative moves: a claim supported by data, warrants linking the data to the claim, backing for support, qualifiers indicating the claim's strength, and rebuttals addressing potential counterarguments. For less-proficient EFL learners, criteria for deep-level features in argumentative writing can be specified further. Adapting research by Siekmann and colleagues (2022), important elements of argumentative essays include an introductory sentence leading to the essay's topic, a thesis statement expressing the opinion, a transitional sentence previewing the main body, counterarguments and rebuttals, multiple pro-arguments with supporting evidence, a summary of arguments, a rephrased thesis, and a closing sentence with a call to action or other concluding strategies. These elements should be effectively linked and organized with appropriate paragraph breaks.

Writing self-efficacy and writing enjoyment

Students' struggles with writing may also stem from ensuing affective and motivational states (Bruning & Kauffman, 2016), given that writing demands substantial motivation to sustain engagement (Graham, 2018), which may also explain inconsistencies in studies examining the effects of formative feedback (Shute, 2008). Key factors influencing persistence and effort in writing include writing self-efficacy (Klassen, 2002; Usher & Pajares, 2008; see also the overview by Kormos, 2012) and writing enjoyment (Tahmouresi & Papi, 2021; Zumbunn et al., 2019), both of which impact writing quality and quantity, and may themselves be influenced by writing experiences.

Self-efficacy is the belief in one's ability to complete a specific task (Bandura, 1977). In writing, this may involve capability beliefs regarding communication via writing and mastering writing tasks (Busse et al., 2023; Klassen, 2002; Pajares, 2003). Generally, self-efficacy related to specific writing tasks is positively associated with writing quality (e.g., Busse et al., 2023; see also the systematic review by Camacho et al., 2021), which can also be observed in EFL writers (e.g., Hetthong & Teo, 2013; Teng & Wang, 2023; Teng et al., 2018). The meta-regression analysis by Sun and colleagues (2021) further suggests that the relationship between self-efficacy and writing achievement is stronger in language learners ($r=0.441$) than in L1 learners ($r=0.233$). However, included studies again focused on higher education, leaving a gap regarding secondary students. Writing in a less-familiar language is often linked to low self-efficacy (Mills, 2014; Pajares, 2003), exacerbated by fear of making errors when using an FL in challenging contexts (Cheng, 2004). This can reduce confidence and persistence, complicating the writing process and thereby limiting students' chances of experiencing mastery. However, despite lower proficiency of multilingual writers, Busse and colleagues (2023) found comparable levels of self-efficacy between L1 and multilingual writers. These contradictory findings may result from students over- or underestimating their abilities, affecting the alignment between self-efficacy and performance (Anastasiou & Michail, 2013;

Busse et al., 2023). Because self-efficacy interacts with other success factors like self-regulation (Zimmerman & Risemberg, 1997; see also Anastasiou & Michail, 2013), overestimating abilities—known as the illusion of competence (see also Kruger & Dunning, 1999)—may be less harmful to writing development than underestimating.

In general, research shows that enjoyment of the target language correlates with increased willingness to use it (Khajavy et al., 2018; Lee, 2022) and writing enjoyment, or pleasure from writing activities, is an important affective-motivational factor influencing writing (see e.g., Tahmouresi & Papi, 2021; Zumbunn et al., 2019). However, existing cross-sectional studies show mixed results on the relationship between writing achievement and writing enjoyment. For instance, Tahmouresi and Papi (2021) found no link between writing enjoyment and writing achievement among tertiary EFL writers ($N=85$), while Li and colleagues (2023) identified small to medium positive correlations among secondary EFL learners ($N=1036$).

Although research on writing enjoyment is limited, existing evidence further suggests that university students who enjoy writing more also tend to use more writing strategies (Zhao et al., 2023). Writing enjoyment also seems to be positively linked to self-regulatory behavior, which, in turn, may affect writing consistency (Zumbunn et al., 2019).

When aiming to support affective-motivational variables, one may assume that mastery experiences in writing can enhance self-efficacy (Bandura, 1977) and possibly also writing enjoyment. Although intervention research is limited, findings from feedback interventions in higher education (Duijnhouwer et al., 2010; Zarrinabadi & Rezaadeh, 2020), demonstrates that responses to writing should provide information on progress toward the learning goal (*feed back* see Hattie & Timperley, 2007) to foster self-efficacy beliefs, likely because it makes mastery experiences visible. Similarly, research with secondary students suggests that progress feedback improves writing self-efficacy (Schunk & Swartz, 1993). While writing instruction alone may not significantly enhance enjoyment (Hidi et al., 2002), it is possible that more extensive writing support, such as feedback, may be required to influence these variables, thus further research is necessary.

In summary, while correlational studies have demonstrated positive relationships between writing achievement and affective-motivational variables, there is lack of research involving secondary students and limited evidence from intervention studies investigating the effect of feedback on self-efficacy and enjoyment.

Feedback perceptions

For feedback to effectively assist learners in navigating the writing process and managing affective-motivational states, it must be designed for uptake, meaning learners' perspectives should inform its design to ensure positive perceptions (Carless & Boud, 2018). Such positive perceptions can promote self-regulation (e.g., He et al., 2023), possibly because feedback perceptions mediate the relationship between affective-motivational factors (e.g., Zumbunn et al., 2016).

A systematic review by Berry and colleagues (2025) highlights that students typically appreciate written teacher comments, though positive feedback perceptions tend to decrease with age. Unspecific feedback, overly focused on error correction, or lacking positive reinforcement can undermine motivation (Busse, 2013; see also the systematic reviews by Ekholm et al., 2018; Winstone et al., 2017). To avoid such pitfalls, our intervention made feedback criteria transparent, emphasized deep-level feedback, and incorporated selective error correction and positive reinforcement. However, we have yet to explore less-proficient writers' perceptions of feedback methods like rubrics compared to in-text comments.

Feedback adhering to quality criteria outlined by Hattie and Timperley (2007) is generally well-received by students (Brooks et al., 2019). In addition, rubrics may help less-proficient writers by providing a consistent framework that clarifies expectations. Yet, rubrics require a high individual engagement, which may alienate struggling students who feel unsupported in interpreting criteria independently. In contrast, in-text comments provide more explicit guidance, reducing the burden of independently identifying solutions, which may benefit at-risk writers. However, if in-text comments are too critical or overwhelming, they could negatively impact motivation. Both rubrics and in-text comments have the potential to support at-risk writers, but their effectiveness likely depends on individual needs and the feedback context. Exploring secondary students' perceptions of these methods is, therefore, important.

Aims and research questions

Our aim is to explore the differential effects of formative feedback in deep-level features and affective-motivational variables among at-risk and developing writers. Specifically, we will investigate the following four research questions (RQs):

- RQ1: To what extent does the type of formative feedback (experimental group [EG] 1: rubric + exemplar; EG2: in-text comments; EG3: combination of rubric + exemplar and in-text comments) influence the deep-level features of argumentative essays in writers with different achievement levels?
- RQ2: To what extent does the type of formative feedback affect writing self-efficacy in writers with different achievement levels?
- RQ3: To what extent does the type of formative feedback affect writing enjoyment in writers with different achievement levels?
- RQ4: To what extent does the type of formative feedback affect feedback perceptions in writers with different achievement levels?

Method

Design

We conducted a randomized controlled intervention study with a pre-, post-, and follow-up design. For the present study, we measured the effect of a writing treatment and different feedback methods on writing development during the intervention by exploring changes in deep-level features in the EGs in four argumentative essays compared to a control group (CG) that only received the writing treatment without feedback. For RQ1, we examine changes in the essays' deep-level features, including four conditions and two achievement levels over four measurement points during the intervention ($4 \times 2 \times 4$ factorial design). For RQs 2 and 3, we investigate writing self-efficacy and writing enjoyment, covering four conditions and two achievement levels over two measurement points (pre- and posttest; $4 \times 2 \times 2$ factorial design), and for RQ4, we explore feedback perceptions, comprising four conditions and two achievement levels at one measurement point ($4 \times 2 \times 1$ factorial design).

Participants

The larger project in which the present study is embedded was conducted with a total of 294 EFL students in Year 9 at comprehensive schools with lower social indices in North Rhine-Westphalia, Germany. For the present study investigating the essays students wrote during the intervention, we did not include the $n=92$ students from the second control group (CG2) as this group did not receive any intervention. Instead, they only participated in pre- and post-testing and continued with business-as-usual, i.e., regular English lessons (approximately 3 h per week), where writing was taught by regular teachers. Thus, the sample for the present study comprised $N=202$ students ($M_{\text{age}} = 14.82$ years, $SD_{\text{age}} = 0.61$, 51.0% girls). We split students in all intervention groups (EGs and CG1) with a median split,¹ based on their scores in overall writing quality in English (holistically rated via comparative judgment using the software Comproved®, see also Peltzer et al., 2024) at the pretest. Although all students may be characterized as beginner writers of English, we will refer to the resulting achievement-level groups as developing ($n=101$; median and up) and at-risk writers ($n=101$; below median) for the following reasons: Firstly, at-risk writers on average achieved less than 15% of the theoretical maximum score in the English writing pretest, which was aligned with curricular expectations for Year 9. This group is thus in serious danger of falling short of meeting curricular expectations, requiring them to achieve CEFR proficiency level B1² at the end of Year 10.

¹ The median itself was added to the group of higher-achieving writers.

² At CEFR level B1, writers are expected to produce clear, connected texts on familiar topics, organizing shorter elements into a coherent sequence. They should also be able to write about topics of personal interest, presenting advantages and disadvantages, and expressing their opinions with basic justification (see Council of Europe, 2018).

Secondly, this group also had significantly lower grades in English and German and scored lower in the general cognitive abilities test.

Writing treatment

As the foundation of our intervention, we developed a learning unit on argumentative writing that combined a process-oriented approach with genre pedagogy. Each session focused on key writing and revising components. In the first session (Writing 1), students wrote an essay on student exchanges abroad (Prompt: '*Student exchanges are not worth the trouble.*'), followed by a revision in the next session (Revision 1) based on the content of the learning unit (in CG1) and additional formative feedback in the EGs. EG1 and EG3 also completed an activity identifying rubric criteria in an exemplar. This pattern was repeated in sessions 3 and 4 with a new writing prompt on a similar topic (Prompt: '*All students who want to go abroad should attend a preparatory class.*'), followed by another revision, in the EGs again based on feedback. EG1 and EG3 completed another exemplar activity, filling in blanks with linking words. Procedural writing activities, such as pre-writing tasks, were paired with genre-oriented activities, all contextualized within a continuous storyline using custom-developed comics (see Lira Lorca et al., 2024).

Instruments for independent variable: formative feedback on deep-level features

Students in the EGs received the writing treatment (learning unit) and additional written formative feedback in the sessions following their writing tasks, specifically in sessions 2 and 4, approximately five days after writing the respective first drafts. All feedback methods were designed to address the same aspects of the essay, met the quality criteria outlined by Hattie and Timperley (2007), and were individualized to the single student's essay (see also Fleckenstein et al., 2024). Students in the control condition (CG1) only received the learning unit and were asked to self-revise based on its contents.

Rubric

We designed an analytic rubric outlining the expected essay criteria along with checkboxes differentiating three degrees of criterion fulfillment ("yes", "partly", "not yet") per criterion (for an overview of rubric implementation in this study, see also Appendix 1). The rubric comprised twelve criteria in total: an *introductory sentence*, which leads to the essay's topic; a *thesis statement*, in which students express their opinion about the topic; a *transitional sentence*, which announces what will be following in the main body; a *counterargument* in opposition to the thesis statement; a *rebuttal*, which challenges the counterargument; at least two *pro-arguments* in favor of the thesis statement; *support* for each pro-argument, such as an example, an explanation, an expert's opinion, or facts; a *summary* of the main arguments; a *rephrasing of the thesis statement*; and a *closing sentence*, such as a call to action, a piece of advice, a prediction, or a compromise. In addition, the rubric

covered the inclusion of *paragraph breaks* that separate the introduction, main body, and conclusion sections, as well as *linking words* that connect different ideas. After each criterion, abbreviations were added in square brackets, for example “[TH]” for *thesis statement*. We created an analytic rating scheme precisely delineating what no, partial, or full fulfillment of each criterion would look like (see Peltzer et al., 2025).

Exemplar

For each of the writing prompts, we created an exemplar illustrating ideal task fulfillment, aligning with the rubric criteria. Square brackets with abbreviations for each respective criterion were included in the corresponding sections of the exemplar to indicate what part of the exemplar showed which criterion of the rubric. Both exemplars were comparable in length with 210 and 243 words each.

In-text comments

For in-text comments, we created a pool of 70 pre-prepared comments for three degrees of criterion fulfillment (“yes”, “partly”, “not yet”) corresponding with the rating scheme (see Peltzer et al., 2025) informing the rubric to ensure comments aligned with rubric criteria and to maintain consistency and feedback quality. Four more comments were added to this pool during the intervention as needed.³ A decision-making flowchart was also developed to prioritize decisive elements like *thesis statements* and *arguments*, avoiding overwhelming students with too many comments. In-text comments were pre-written digitally in Adobe Acrobat®.

Instruments for dependent variables

Pretest writing quality (holistically measured)

At the pretest, students in all groups wrote an argumentative essay where they were asked to discuss the statement “The internet makes traveling unnecessary.”. Overall writing quality in these essays was measured holistically. Drawing on the software Comproved® (see Goossens & De Maeyer, 2018), seven university students majoring in teaching English conducted a comparative judgment of all pooled pre- and posttest essays for this prompt. Ratings showed good internal consistency (SSR=0.81) (see also Peltzer et al., 2024a). In addition, we conducted an analytic rating of two benchmark essays, which were converted to a 0–100 scale,

³ These instances included the use of incomplete sentences, where parts of the essay were presented as bullet points instead of complete sentences. There was also improper argument placement, where new arguments were introduced in the introduction or conclusion, rather than being developed in the main body of the essay. Additionally, there was improper summary placement, where the summary of the main ideas was not placed at the beginning of the conclusion as required. Finally, there was unclear thesis rephrasing, where the rephrasing of the thesis statement lacked clarity and failed to adequately explain the topic of the original thesis statement.

and Comproved® generated percentage scores for the remaining essays as indicators of writing quality. This measurement was used to conduct the median split in this study.

Deep-level features in essays during intervention (analytically measured)

During the intervention, students wrote and subsequently revised essays on two prompts. During the first session (Writing 1), students wrote an essay discussing the statement “Student exchanges are not worth the trouble.”, which was revised in the adjacent session (Revision 1). In the third session (Writing 2), students wrote a new essay on the prompt “All students who want to go abroad should attend a preparatory class.”, which was also revised in the next session (Revision 2), resulting in four essays per student in the intervention groups (EGs and CG1), which we then analyzed for their deep-level features. We measured deep-level features in the 808 essays that students wrote during the intervention using an analytic rating scheme adapted from Siekmann and colleagues (2022). Focusing on the coherence and structure of the target text type (argumentative essay), the scheme differentiated 12 criteria, covering very genre-specific elements such as the inclusion of a thesis statement and arguments, but also more general structural elements, such as the division of the text into introduction, main body, and conclusion. We awarded points on a three-point scale, ranging from zero to two, for each criterion. Thus, a maximum of 24 points was awarded per essay. The full rating scheme can be found in Peltzer and colleagues (2025). The two doctoral students who conducted the intervention also assessed the deep-level features, making independent assessments first, which were then discussed until consensus was reached in one joint score (see also Trace et al., 2016). A random 20% sample of the essays, equally from each measurement point, were re-evaluated by a different rater unaware of the feedback condition and measurement point ($ICC = .980$).

Writing self-efficacy

Writing self-efficacy was measured at pre- (T1) and posttest (T2) using questionnaires with a six-point Likert scale. To assess writing self-efficacy, we adapted a scale from Bruning and colleagues (2013); see also Siekmann et al., (2023) to fit the content of our intervention using nine items, with satisfactory Cronbach’s alpha values above .8 (see Appendix 2).

Writing enjoyment

Writing enjoyment was also measured at pre- (T1) and posttest (T2) using questionnaires with a six-point Likert scale. To assess writing enjoyment, we drew on a scale adapted from Zumbrunn and colleagues (2019; see also Busse et al., 2020) using four items, with satisfactory Cronbach’s alpha values above .8 (see Appendix 2).

Feedback perceptions

We measured students' perceptions of feedback at two measurement points during the intervention, using six-point Likert scales. One scale examined the perceived quality of feedback according to Hattie and Timperley's (2007) criteria using seven items, while another scale assessed the feedback's motivational impact with three items (see also Siekmann et al., 2023). For the present study, we aggregated scores from the two measurement points during the intervention for each scale separately. Additionally, at T2, we requested a summative feedback evaluation from students, via four items in the questionnaire, also measured on a six-point Likert scale. All perception scales had satisfactory Cronbach's alpha values above .8 (see Appendix 3).

Procedure

Our study adhered to the ethical guidelines of the German Research Foundation and received university ethics approval. We obtained informed written consent from all participants. To recruit participants, we contacted all comprehensive schools within a 100 km radius of the University of Münster and eight schools agreed to participate. We conducted a pretest with all students, who were then randomly assigned to one of three EGs receiving different types of formative feedback (EG1: rubric + exemplar, $n=55$; EG2: in-text comments, $n=55$; EG3: combination of rubric + exemplar and in-text comments, $n=42$)⁴; or a control group (CG1, $n=50$)⁵ that received the learning unit without additional written feedback. After randomly assigning students to the EGs and CG1, we implemented the intervention with all EGs and CG1 (see Fig. 1). Our researcher-led intervention included a writing treatment—a 4.5-h learning unit on argumentative writing across four sessions—and formative feedback for the EGs. Students in EG1 received a feedback sheet with a rubric and exemplar (see Appendix 3). Two doctoral students (i.e., the first and second authors) filled out the checkboxes ("yes", "partly", "not yet") to assess each criterion. In addition, a small section of the sheet was reserved for selected surface-level errors (spelling, word choice, grammar, tenses) that were corrected selectively (Bitchener & Knoch, 2010; Ferris, 2011). EG2 received in-text comments in the margins of their essays. Comments on structure and coherence were placed in the left margin. For example, if the criterion *pro-argument* (There are at least two pro-arguments, which support your thesis statement) was not met at all, we wrote the comment 'In your main body, please add at least two reasons (pro-arguments) that support your opinion (thesis statement), e.g., 'One reason is that ...' in the essay margin. If the criterion was partially fulfilled, we provided feedback tailored to the scenario: 'Please

⁴ Group sizes varied due to the lack of consent from some students and the absence of some participants during lessons. Only complete data sets were analyzed.

⁵ We also collected data from an additional control group 2 (CG2) that participated in pre-, post-, and follow-up testing but did not receive any intervention (see also Peltzer et al., 2024). Since we are not reporting any data from this CG2 in the present study, we did not include it in the description of the sample.

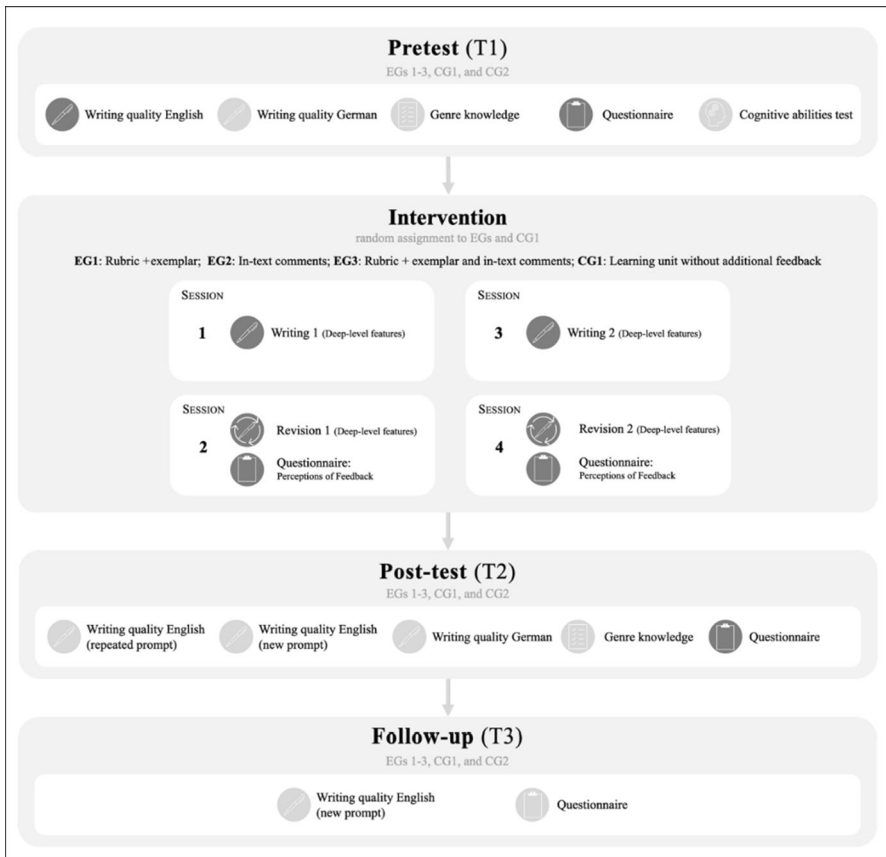


Fig. 1 Overview of study design and data collected *Note:* Dark grey icons indicate data used for this study

add (at least) one more pro-argument', 'Please change this pro-argument so that it supports your thesis statement', or 'Please change this pro-argument so that it supports your thesis statement and add (at least) one more pro-argument', or 'Please move your (pro-)arguments to your main body'. If the criterion for pro-arguments was fully met, we added 'Good pro-argument!' to each pro-argument. Comments related to selective surface-level error correction were placed in the right margin. In-text comments were pre-written digitally in Adobe Acrobat®. Students' essays were scanned, and in-text comments were selected individually in response to each essay and then placed on the PDF files using the 'stamp' function. Then, we printed out the commented scan and handed it back to the students. Students in EG3 received both the rubric + exemplar and in-text comments, allowing students to choose which feedback to prioritize during their revision process and how to engage with each form of feedback, i.e., students in EG3 were not instructed on which feedback to prioritize.

The two doctoral students (i.e., the first and second authors) taught all sessions of the learning unit based on a standardized PowerPoint presentation aligned with a detailed script and a structured protocol and implemented the formative teacher feedback (EGs 1–3). Each participating student also received standardized workbooks containing all classroom materials used in the sessions. While one researcher taught, the other took detailed notes following the structured protocol. This documentation covered every step of the instruction, including specific explanations of argumentative elements, task guidelines and instructions, and other relevant aspects of the learning unit. All slides of the PowerPoint presentation were shown and all instructional content (100% of the intervention) was fully implemented in every group. Each session was scheduled to last either 45 or 90 min, though the actual duration varied slightly by a few minutes due to factors such as group engagement or occasional disruptions. Nonetheless, the time allocated for writing and revision was consistent across all groups, with 15 minutes designated for writing and 20 minutes for revision. To prevent cross-group interference, students were separated during feedback and revision phases. The intervention concluded with a posttest and a follow-up twelve weeks later.

Statistical analyses

We first analyzed pretest data to ensure group comparability, finding no differences in overall writing quality (holistically rated via comparative judgment using the software Comproved®) and cognitive abilities⁶ (see also Peltzer et al., 2024a) in the conditions, with holistic writing quality generally being low. Conditions (EGs and CG1) were also comparable at the first intervention measurement (Writing 1) regarding deep-level features in the essays ($F[3, 189]=0.100, p=0.960, \eta_p^2=0.002$). After performing the median split based on pretest writing quality in English, we found no uneven distributions of developing and at-risk writers in the conditions, i.e., all conditions contained comparable numbers of at-risk ($F[3, 100]=0.671, p=0.572, \eta_p^2=0.020$) and developing writers ($F[3, 100]=0.879, p=.455, \eta_p^2=.026$). Correlations for all variables are detailed separately for at-risk writers in Appendix 4 and for developing writers in Appendix 5.

Further analyses showed that at-risk writers also had significantly lower grades than developing writers in English ($t[194]=7.953, p<.001, d=1.136, 95\%—CI [0.833; 1.437]$) and in German ($t[194]=3.048, p<.001, d=0.435, 95\%—CI [0.152; 0.718]$), i.e., in subjects where writing plays a crucial role in assessment. They also scored lower on the general cognitive abilities test ($t[195.718]=-4.213, p<.001, d=-0.593, 95\%—CI [-0.874; -0.310]$).

To address our research questions, we used repeated measures ANOVAs to analyze potential interactions between time, condition, and achievement level. If an interaction was found, we further examined at-risk and developing writers separately

⁶ We examined students' general cognitive abilities using a subtest on figural reasoning (KFT 4-12+R, subtest B, derived from Heller and Perleth, 2000).

to identify sources of interaction. If no interaction was detected, we analyzed achievement-level groups across conditions and assessed the main effects using repeated measures ANOVAs. For significant effects related to achievement level alone, we conducted additional t-tests. All statistical requirements were met unless noted, with adjustments reported.

To interpret our data, we used the benchmarks suggested by Cohen (1988) for t-tests of small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d < 0.8$), and large ($0.8 \leq d$) effects, and for ANOVA of small ($0.01 \leq \eta_p^2 < 0.06$), medium ($0.06 \leq \eta_p^2 < 0.14$), and large ($0.14 \leq \eta_p^2$) effects. Data analyses were conducted with IBM SPSS Statistics v.29.

Results

RQ1: Effects of feedback types on deep-level features (total scores) in essays of writers with different achievement levels

To examine how different feedback types affected deep-level features in at-risk and developing writers, we analyzed the data by condition and achievement level. At the first measurement point (Writing 1), students in neither condition nor achievement level reached half of the theoretical maximum (see Table 1). Results indicated a significant difference between achievement-level groups in implementing deep-level features like structure and coherence (Writing 1: $t[200] = -5.793$, $p < .001$, $d = -0.815$, 95%—CI [-1.101, -0.527]). By the last measurement point (Revision 2), all conditions and achievement levels showed improvement, with EG1 making the largest descriptive progress. At-risk writers in EG2 and EG3 made similar progress to those in CG1 who received no additional feedback (see Fig. 2).

A repeated measures ANOVA with condition and achievement level as between-subject factors showed a significant interaction of time, condition, and achievement level ($F[9, 582] = 2.373$, $p = .012$, $\eta_p^2 = .035$), and a significant interaction effect for time and condition ($F[9, 582] = 7.172$, $p < .001$, $\eta_p^2 = .100$). However, interactions between achievement level and condition ($F[3, 194] = 0.375$, $p = .770$, $\eta_p^2 = .006$) and between achievement level and time ($F[3, 582] = 1.091$, $p = .352$, $\eta_p^2 = .006$) were not significant. Significant main effects were found for time ($F[3, 582] = 120.733$, $p < .001$, $\eta_p^2 = .384$) and achievement level ($F[1, 194] = 50.416$, $p < .001$, $\eta_p^2 = .206$), but not for condition ($F[3, 194] = 0.376$, $p = .770$, $\eta_p^2 = .006$).

For a closer investigation, we ran separate post-hoc tests for at-risk and developing writers. For at-risk writers, repeated measures ANOVA showed a significant time effect ($F[3, 291] = 60.043$, $p < .001$, $\eta_p^2 = .382$), no significant group effect (condition) ($F[1, 97] = 2.495$, $p = .064$, $\eta_p^2 = .072$), but a significant interaction effect ($F[9, 291] = 4.502$, $p < .001$, $\eta_p^2 = 0.122$). Post-hoc analysis revealed that EG1 (rubric + exemplar) significantly ($p = .039$) outperformed CG1 ($M_{\text{Diff}} = 2.625$, 95%—CI [0.091; 5.159]). Among developing writers, repeated measures ANOVA revealed a significant time effect ($F[3, 291] = 61.726$, $p < .001$, $\eta_p^2 = .389$), no significant

Table 1 Scores for deep-level features in essays during the intervention among at-risk and developing writers: Means and standard deviations

	EG1		EG2		CG1		EG3		All EGs and CG1	
	At-risk writers <i>n</i> = 26 <i>M</i> (<i>SD</i>)	Developing writers <i>n</i> = 29 <i>M</i> (<i>SD</i>)	At-risk writers <i>n</i> = 30 <i>M</i> (<i>SD</i>)	Developing writers <i>n</i> = 25 <i>M</i> (<i>SD</i>)	At-risk writers <i>n</i> = 19 <i>M</i> (<i>SD</i>)	Developing writers <i>n</i> = 23 <i>M</i> (<i>SD</i>)	At-risk writers <i>n</i> = 26 <i>M</i> (<i>SD</i>)	Developing writers <i>n</i> = 24 <i>M</i> (<i>SD</i>)	At-risk writer <i>n</i> = 101 <i>M</i> (<i>SD</i>)	Developing writers <i>n</i> = 101 <i>M</i> (<i>SD</i>)
Deep-level features (Writing 1)	6.31 (3.68)	8.76 (4.87)	6.60 (3.53)	9.48 (4.46)	6.74 (3.44)	9.13 (3.51)	5.31 (3.07)	10.42 (4.39)	6.22 (3.44)	9.42 (4.35)
Deep-level features (Revision 1)	9.85 (5.09)	13.65 (6.26)	10.27 (4.85)	15.32 (5.16)	10.37 (5.24)	11.83 (6.09)	8.77 (4.02)	14.37 (4.26)	9.79 (4.76)	13.82 (5.59)
Deep-level features (Writing 2)	10.96 (5.54)	14.38 (5.70)	10.33 (4.34)	13.44 (6.12)	7.53 (4.38)	14.04 (5.16)	7.61 (3.70)	11.54 (4.18)	9.27 (4.73)	13.40 (5.40)
Deep-level features (Revision 2)	15.04 (4.33)	18.17 (4.78)	10.57 (4.83)	14.36 (6.13)	11.58 (4.45)	16.87 (5.50)	9.96 (3.86)	14.46 (4.84)	11.75 (4.78)	16.05 (5.50)
Δ _{Revision 2 - Writing 1}	8.73	9.41	3.97	4.88	4.84	7.74	4.83	4.04	5.53	6.63

Means and standard deviations based on scores for deep-level features (structure and coherence) in essays. Theoretical minimum = 0; theoretical maximum = 24.

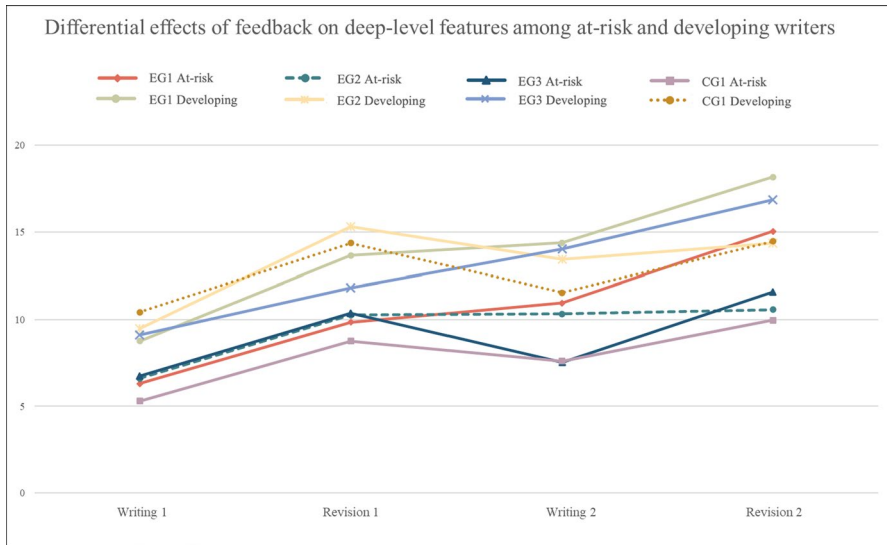


Fig. 2 Development of deep-level features in essays during the intervention among at-risk and developing writers: Mean scores

group (condition) effect ($F[3, 97]=0.296, p=.828, \eta_p^2=.009$), but a significant interaction effect ($F[9, 291]=4.923, p<.001, \eta_p^2=.132$).

RQ2: Effects of feedback types on writing self-efficacy in writers with different achievement levels

Students at both achievement levels generally felt self-efficacious, with scores in the middle to upper range. However, T1 scores were significantly lower among at-risk writers ($t[194]=-4.345, p<.001, d=-0.621, 95\% \text{—} \text{CI} [-0.907, 0.333]$) (see Table 2) and remained lower than those of developing writers at T2. To examine the differential effects of feedback types on writing self-efficacy, we first analyzed the data by condition and achievement level. A repeated measures ANOVA with condition and achievement level as between-subject factors showed no significant interaction of time, condition, and achievement level ($F[3, 181]=0.246, p=.864, \eta_p^2=.004$). Since no interaction was found, we further examined the interactions between time and condition ($F[3, 181]=0.928, p=.428, \eta_p^2=.015$) and between time and achievement level ($F[1, 181]=0.379, p=.539, \eta_p^2=.002$), neither of which were significant. However, we did find a significant main effect of achievement level ($F[1, 181]=21.137, p<.001, \eta_p^2=.105$), while neither the main effect of time ($F[1, 181]=2.140, p=.145, \eta_p^2=.012$) nor of condition ($F[3, 181]=0.813, p=.488, \eta_p^2=.013$) reached significance. This indicated that self-efficacy remained relatively stable over time in both achievement-level groups.

Table 2 Writing self-efficacy among at-risk and developing writers in EGs and CG1 (*n* = 202): Means and standard deviations

	EG1		EG2		EG3		CG1		All EGs and CG1	
	At-risk writers <i>n</i> = 26	Developing writers <i>n</i> = 29	At-risk writers <i>n</i> = 30	Developing writers <i>n</i> = 25	At-risk writers <i>n</i> = 19	Developing writers <i>n</i> = 23	At-risk writers <i>n</i> = 26	Developing writers <i>n</i> = 24	All at-risk writer <i>n</i> = 101	All developing writer <i>n</i> = 101
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Writing self-efficacy (T1)	3.82 (1.13)	4.34 (0.75)	3.80 (1.01)	4.43 (0.67)	3.99 (1.42)	4.76 (0.95)	4.10 (1.04)	4.61 (0.65)	3.92 (1.12)	4.52 (0.76)
Writing self-efficacy (T2)	4.01 (0.79)	4.55 (0.72)	3.99 (1.00)	4.50 (0.95)	4.18 (0.93)	4.72 (1.00)	4.02 (0.94)	4.56 (0.74)	4.04 (0.91)	4.58 (0.85)
Δ T2-T1	0.19	0.21	0.19	0.07	0.19	- 0.04	- 0.08	- 0.05	0.12	0.06

Means and standard deviations based on a six-point Likert scale ranging from 1 (not true at all) to 6 (very much true).

RQ3: Effects of feedback types on writing enjoyment in writers with different achievement levels

At T1, writing enjoyment scores were in the middle to high range for both achievement-level groups, with at-risk writers reporting lower writing enjoyment than developing writers ($t[197] = -5.266$, $p < .001$, $d = -0.747$, 95%—CI [- .033, -0.458]) (see Table 3). Scores remained stable among at-risk writers, while scores slightly declined among developing writers. To investigate the differential effects of feedback types on writing enjoyment, we again first analyzed the data by condition and achievement level. A repeated measures ANOVA with condition and achievement level as between-subject factors showed no significant interaction of time, condition, and achievement level ($F[3, 188] = 1.087$, $p = .356$, $\eta_p^2 = .017$). Since no significant interaction was found, we examined interactions between time and condition ($F[3, 188] = 0.324$, $p = .808$, $\eta_p^2 = .005$), which was not significant, and between time and achievement level ($F[1, 188] = 6.260$, $p = .013$, $\eta_p^2 = .032$), which was significant. Additionally, there was a significant main effect of achievement level ($F[1, 188] = 18.545$, $p < .001$, $\eta_p^2 = .090$), whereas the main effect of time was not significant ($F[1, 188] = 1.445$, $p = .229$, $\eta_p^2 = .008$).

Finally, pairwise comparisons with adjusted alpha level ($\alpha = 0.05/2 = 0.025$) revealed that writing enjoyment among at-risk writers remained stable ($t[84] = 0.104$, $p = .917$, $d = 0.011$, 95%—CI [- 0.201, 0.224]), while losses among developing writers reached significance ($t[96] = -2.617$, $p = .010$, $d = -0.266$, 95%—CI [-0.468, 0.063]).

RQ4: Feedback perceptions in writers with different achievement levels

Feedback was perceived positively by the students in the EGs, as indicated by values in the upper range of the scales both during the intervention and at T2 on both achievement levels (see Table 4). Feedback quality was perceived more positively than the motivational impact of feedback both among at-risk writers ($t[66] = 10.790$, $p < .001$, $d = 1.318$, 95%—CI [0.987, 1.644]) and developing writers ($t[67] = 9.423$, $p < .001$, $d = 1.143$, 95%—CI [0.834, 1.446]).

To investigate how at-risk and developing writers perceived the different feedback conditions, we conducted one-way ANOVAs with adjusted alpha level ($\alpha = 0.05/2 = 0.025$) with condition as a between-subject factor for each achievement-level group separately. Among at-risk writers, we found no differences between conditions in terms of perceived quality of feedback ($F[2, 64] = 1.320$, $p = .274$, $\eta_p^2 = .040$), perceived motivational impact of feedback ($F[2, 66] = 1.597$, $p = .210$, $\eta_p^2 = .046$), or summative feedback evaluation ($F[2, 71] = 0.547$, $p = .581$, $\eta_p^2 = .015$). Similarly, among developing writers, we found no significant differences between conditions on either scale, neither regarding perceived quality of feedback

Table 3 Writing enjoyment among at-risk and developing writers in EGs and CG1 (*n* = 202); Means and standard deviations

	EG1		EG2		EG3		CG1		All EGs and CG1	
	At-risk writers <i>n</i> = 26	Developing writers <i>n</i> = 29	At-risk writers <i>n</i> = 30	Developing writers <i>n</i> = 25	At-risk writers <i>n</i> = 19	Developing writers <i>n</i> = 23	At-risk writers <i>n</i> = 26	Developing writers <i>n</i> = 24	All at-risk writers <i>n</i> = 101	All developing writers <i>n</i> = 101
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Writing enjoyment (T1)	3.39 (1.38)	3.94 (1.07)	3.11 (1.49)	4.19 (1.32)	3.53 (1.39)	4.66 (0.86)	3.20 (1.36)	4.20 (1.10)	3.29 (1.40)	4.22 (1.12)
Writing enjoyment (T2)	3.49 (1.30)	3.89 (1.10)	3.34 (1.48)	3.78 (1.54)	3.71 (1.55)	4.22 (1.50)	3.07 (1.37)	4.01 (1.10)	3.38 (1.42)	3.96 (1.30)
Δ T2-T1	0.10	-0.05	0.23	-0.41	-0.18	-0.44	-0.13	-0.19	0.09	-0.26

Means and standard deviations based on a six-point Likert scale ranging from 1 (not true at all) to 6 (very much true).

Table 4 Feedback perceptions in the EGs ($n = 152$): Means and standard deviations

	EG1		EG2		EG3		All EGs	
	At-risk writers $n = 26$	Developing writers $n = 29$	At-risk writers $n = 30$	Developing writers $n = 25$	At-risk writers $n = 19$	Developing writers $n = 23$	At-risk writers $n = 75$	Developing writers $n = 75$
	$M (SD)$	$M (SD)$	$M (SD)$	$M (SD)$	$M (SD)$	$M (SD)$	$M (SD)$	$M (SD)$
Perceived quality of feedback (during intervention)	4.79 (0.69)	5.00 (0.62)	5.12 (0.68)	5.20 (0.78)	5.11 (0.95)	5.24 (0.65)	5.02 (0.77)	5.14 (0.69)
Perceived motivational impact of feedback (during intervention)	3.42 (1.07)	3.69 (1.13)	3.95 (1.30)	3.96 (1.47)	4.04 (1.24)	4.06 (1.37)	3.81 (1.23)	3.90 (1.32)
Summative evaluation of feedback (T2)	4.58 (1.12)	4.81 (1.19)	4.88 (1.09)	4.78 (1.27)	4.72 (0.97)	4.65 (1.36)	4.73 (1.06)	4.75 (1.25)

Means and standard deviations based on a six-point Likert scale ranging from 1 (not true at all) to 6 (very much true).

($F[2, 67]=0.827, p=.442, \eta_p^2=.024$), nor regarding perceived motivational impact of feedback ($F[2, 67]=0.486, p=.617, \eta_p^2=.014$), nor in the summative evaluation of feedback ($F[2, 71]=0.110, p=.896, \eta_p^2=.003$). When considering feedback perceptions regardless of condition we again found no significant differences between at-risk and developing writers in perceived quality of feedback ($t[135]=-0.979, p=.329, d=-0.167, 95\%—CI [-0.503, 0.169]$), perceived motivational impact of feedback ($t[137]=-0.376, p=.707, d=-0.064, 95\%—CI [-0.396, 0.269]$), or summative evaluation of feedback ($t[146]=-0.108, p=.916, d=-0.017, 95\%—CI [-0.340, 0.305]$).

Discussion

Our feedback intervention suggests that at-risk writers exhibited larger deficits in deep-level features such as structure and coherence compared to developing writers. Regarding *RQI*, we observed differential effects of formative feedback on deep-level features across achievement levels. Interestingly, at-risk writers particularly benefited from rubric+exemplar feedback (EG1), thus aligning with previous research demonstrating the value of rubric-based feedback (Lipnevich et al., 2023; Panadero et al., 2023). In comparison, at-risk writers in EG2 (in-text comments) and EG3 (rubric+exemplar and in-text comments) showed similar progress to at-risk writers in CG1, who received the learning unit without additional feedback.

At-risk writers may experience higher cognitive load than their more proficient peers, which can impede their ability to manage higher-level planning and cohesive writing (Galbraith, 2009). Consequently, reducing cognitive load through automation to facilitate effective problem-solving is particularly important to support language learners. The rubrics may have helped students manage cognitive load by offering a clear, detailed checklist systematically guiding at-risk writers through the revision process. This structured approach may have allowed them to focus on complex elements, like structure and coherence while managing cognitive demands.

The clarity of rubrics ensures writers understand the learning goals (*feed up*) and their current performance (*feed back*) (Lipnevich et al., 2014, 2023; Panadero et al., 2023), making the revision process more manageable and effective thus avoiding surface-level distractions common in struggling writers' revisions (see also Stevenson et al., 2006). While rubrics provide clear *feed up* and *feed back*, *feed forward* (information on next steps required) is more implicit compared to the direct guidance of in-text comments. For instance, in EG1, writers received labels like 'not yet' or 'partly' for specific criteria and had to address issues independently, possibly fostering deeper engagement with the criteria. This is an artifact of the nature of rubrics, instruments created to analyze a sample of performances, versus the nature of feedback comments that are usually written in response to specific pieces of work. For at-risk writers, independently engaging with clear criteria—without overly directive instructions—may be more effective

than in-text comments, fostering greater autonomy. Additionally, the presentation of essay elements in the rubrics provided a repetition of learning content, which may have been especially beneficial for learners with limited prior knowledge. Another component that possibly supported students in EG1 was the exemplar provided alongside the rubrics. By illustrating ideal criterion fulfillment, the exemplar likely helped bridge the gap between rubric criteria (*feed up*) and students' performance (*feed back*), prompting comparisons between students' work and the exemplars (To et al., 2022).

In general, our findings suggest that more feedback does not necessarily improve outcomes for at-risk writers. Results indicate that only the more proficient developing writers benefited from the combination of rubric, exemplar, and in-text comments (EG3), while at-risk writers in EG3 showed progress comparable to the control group. Developing writers may better manage additional cognitive load. These findings underscore the importance of tailored feedback for learners facing high cognitive demands and may also help explain findings by Müller and colleagues (2023): When exploring the relationship between teacher feedback and writing outcomes in mainstream writing classes, the authors found that teachers' *feed forward* was a significant predictor of text quality for students' writing in their first language but not for multilingual learners, suggesting that traditional teacher feedback in the form of in-text comments may not effectively address the needs of struggling writers.

Regarding RQ2, our study found that both groups exhibited relatively high self-efficacy scores (≥ 3.80 on a six-point Likert scale), although at-risk writers demonstrated lower self-efficacy compared to developing writers, which is consistent with previous research indicating that less proficient writers usually have lower writing self-efficacy (e.g., Hetthong & Teo, 2013; Teng & Wang, 2023; Teng et al., 2018). High initial self-efficacy may also explain why self-efficacy did not change during the intervention in either achievement-level group.

The relationship between self-efficacy and writing achievement was notably inconsistent. Learners from both achievement-level groups appeared to struggle with accurately assessing their writing performance, as evidenced by small correlations between self-efficacy and scores in deep-level features in both groups at the first measurement point and moderate correlations among at-risk writers between self-efficacy and holistic writing quality at the pretest. For developing writers, no significant correlations were found between self-efficacy and holistic writing quality at the pretest. These findings suggest a low calibration between self-efficacy and actual performance, consistent with observations in other studies on language learners (Busse et al., 2023; see also Anastasiou & Michail, 2013).

Additionally, as holistic ratings assess overall text quality, including both surface- and deep-level features, while the self-efficacy measure specifically targeted beliefs about deep-level features of argumentative essays, the absence of correlations between self-efficacy and achievement among developing writers at the pretest may indicate that pretest performance differences were driven more by surface-level features than deep-level ones. This interpretation, however, remains tentative, and further research is required to explore the relationship between self-efficacy and writing achievement in less-proficient or struggling writers. If our findings are

replicated in future research, investigation is needed to understand why this is the case. We are not aware of any studies at this point that provide a definitive answer for why there would be a statistically significant association between efficacy and writing quality for lower-achieving students but not higher-achieving students.

Regarding *RQ3*, we also found no differential effects of formative feedback on writing enjoyment, and writing enjoyment remained relatively stable despite a small loss among developing writers, possibly due to the volume of writing assignments not typical in classroom practice (Siekmann & Busse, 2022). However, our sample generally showed medium to high writing enjoyment (≥ 3.1 on a six-point Likert scale), which is encouraging, as enjoyment of the target language is linked to greater willingness among FL learners to use it (Khajavy et al., 2018; Lee, 2022). Enjoying writing may also promote self-regulatory behavior (Zumbrunn et al., 2019), aiding student consistency. It is therefore encouraging that at-risk writers did not experience a loss in writing enjoyment and that all students perceived feedback positively (*RQ4*). In general, our findings support the notion that feedback meeting Hattie and Timperley's (2007) quality criteria is well-received by students (Brooks et al., 2019). It is particularly encouraging that the time-efficient rubric + exemplar feedback was perceived as positively by at-risk writers as in-text comments, despite students probably being more used to the latter.

Learners at all achievement levels rated the perceived quality of feedback significantly higher than its motivational impact, independent of the feedback method. Since the items used to measure the motivational impact of feedback focused on its procedural aspects (e.g., "The feedback made me enjoy revising my essay more"), it is not surprising that lower scores were reported for motivational impact compared to feedback quality, given that revision is challenging for all learners and particularly difficult for our sample of less-proficient writers. Previous studies have emphasized the importance of progress information for positive motivational impact (Duijnhouwer et al., 2010; Schunk & Swartz, 1993). While all experimental groups received feedback on their progress from draft to revision, future studies should prioritize this aspect to enhance the motivational impact of feedback, presenting an opportunity to optimize feedback methods for better student learning.

Limitations and strengths

Our study is notable for its exploration of the differential effects of formative feedback on EFL writers with varying achievement levels, a topic that has received limited attention in existing research. By examining how formative feedback influences both at-risk and developing writers, the study sheds light on the complexities of writing support across varying achievement levels.

The main limitation of this study is the division of the sample by median to differentiate between two achievement-level groups. However, given the four experimental conditions, the sample size was insufficient for more differentiated achievement-level groupings. Future research should explore more fine-grained grouping methods to further refine these findings and deepen our understanding of the role of individual

differences in feedback uptake, including various aspects of prior knowledge as well as other personal characteristics such as language background. Additionally, the study did not isolate the effects of using rubrics or exemplars independently, limiting our ability to make claims about the impact of different feedback components. Lipnevich and colleagues (2023) found that rubrics alone were more effective than when combined with exemplars for L1 writers. Results may differ for less-proficient EFL writers, requiring more research involving students with diverse language backgrounds. Gaining a deeper understanding the cognitive and metacognitive processes involved might be important, here, as feedback effectiveness could depend not only on students' perceptions but also on how they engage with feedback before applying it in writing. Future research could explore these internal processes to gain a more nuanced understanding of feedback's impact. Methods like eye-tracking or follow-up interviews could offer valuable insights into feedback use, though these were beyond the scope of this study and should be considered in future research. Finally, the intervention period, though adequate for detecting changes in deep-level features, may have been too short to significantly impact self-efficacy and writing enjoyment. These motivational factors may require extended time and sustained practice to show measurable improvement. Future research could therefore examine long-term writing development through feedback to better understand its impact on motivation and progress.

Conclusions and implications

In conclusion, our study shows that effective feedback design must consider learners' varying achievement levels. Different types of formative feedback affect EFL writers differently, with at-risk writers showing significant improvement in deep-level features of argumentative essays when given rubric+exemplar feedback. While this type of feedback proved beneficial for all students, our study is the first to show that it has particular benefits for at-risk writers. Rubric+exemplar feedback provides support and clarity during the complex writing process, and may thus help students manage cognitive load (Lipnevich et al., 2023). Students who received in-text comments did not make larger gains in writing than students in the control condition. A combination of both rubric+exemplar and in-text comments, however, seems to be beneficial for higher-achieving developing writers but not for at-risk writers, indicating that more feedback is not necessarily better. While not increasing self-efficacy and enjoyment, all feedback methods were perceived positively. Overall, our study suggests that rubric+exemplar feedback is an effective and time-efficient approach to support struggling writers.

Appendix 1

Instrument to report the characteristics of rubric design and implementation (Version 1.2)

Created by Author

HOW TO SUPPORT AT-RISK WRITERS: DIFFERENTIAL EFFECTS OF FORMATIVE FEEDBACK ON ARGUMENTATIVE WRITING AND MOTIVATION

[Author names anonymized]

Our study investigates:

- Rubrics and scoring accuracy
- Rubrics and academic performance
- Rubrics and students' perceptions
- Rubrics and _____

Describe the characteristics of your rubric intervention study in the table below.

Design		
Category	Description	Our study
1 Rubric presence	Have you included the rubric in the publication as supplementary material?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No. Reason: Click here to add text
2 Assessment criteria	Number of assessment criteria included in the rubric	Twelve individual criteria covering elements of an argumentative essay, plus a small section for selective error correction regarding spelling, grammar, tense, and word choice.
3 Performance levels	How many performance levels are included in the rubric? Also list the headings	Just one. We did not differentiate different levels but only assessed the twelve criteria for each student.
4 Creation	Was the rubric created for this study? If not, please indicate the original source	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
5 Scoring strategy	If the rubric contains an explicit scoring strategy, provide a brief description.	Click here to add text
6 Type	How was the assessment communicated to the students, holistic (i.e., as an overall assessment for all criteria or analytical (i.e., separately for all criteria assessed)?	<input type="checkbox"/> Holistic <input checked="" type="checkbox"/> Analytical
7 Type 2	Was the rubric general (i.e., a general skill such as writing), task-generic (i.e., applicable to several similar tasks) or task-specific (i.e., only applicable to one particular task)	<input type="checkbox"/> General <input checked="" type="checkbox"/> Task-generic <input type="checkbox"/> Task-specific
Implementation		
8 Self-assessment	Was the rubric used for self-assessment?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
9 Self-scoring	Was the rubric used to calculate a self-score?	<input type="checkbox"/> Yes, but the self-score was not included in the final grade. <input type="checkbox"/> Yes, and the self-score represented _% of the final grade. <input checked="" type="checkbox"/> No
10 Peer assessment	Was the rubric used for peer assessment?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No
11 Peer score	Was the rubric used to score a peer?	<input type="checkbox"/> Yes, but the peer score was not included in the final grade.

		<input type="checkbox"/> Yes, and the peer score represented ___% of the final grade. <input checked="" type="checkbox"/> No
12	Feedback	<p>Did the students receive additional feedback about their performance or on how they used the rubric?</p> <p><input checked="" type="checkbox"/> Yes, on both <input type="checkbox"/> Only on their performance <input type="checkbox"/> Only on how they used the rubric <input type="checkbox"/> No</p> <p>If yes, could you describe the additional feedback characteristics? The first and second authors completed the rubric for each student in EG1 and EG3. For the respective first drafts of the two prompts, we indicated criterion fulfillment for the individual essay elements covered by the rubric. For both revisions, we provided students with a progress report, again showing the fulfillment of each rubric criterion. In that, students received a measure of performance but could also deduce information on their progress after using the rubric.</p>
13	Official weight	<p>Did the activity assessed with the rubric count towards the students' grade?</p> <p><input type="checkbox"/> Yes, for a ___% of the total <input checked="" type="checkbox"/> No <input type="checkbox"/> Click here to add text</p>
14	Frequency	<p>How many times was the rubric used? (Once, twice, etc.)</p> <p>Twice</p>
15	Training	<p>Did the participants receive training about the rubric? If yes, describe the training and the specific moment in which they received it.</p> <p>We implemented a learning unit covering four sessions. In the first session, students wrote a draft of their first essay. In the second session, we introduced students to the rubric criteria in an interactive puzzle activity, which we later discussed in class. After that, students received an exemplar in which they had to identify the essay elements covered by the rubric. Then, students received a feedback sheet including both the rubric, filled out for their individual essay, and the exemplar. We explained to students how to interpret the information on the sheet and how it could be used for revision. Students then revised their first draft.</p> <p>In the third session, students wrote another draft for an essay on a new prompt. In the fourth and final session, students did another puzzle activity on linking words and were handed another exemplar matching the new prompt. At this point, the exemplar contained blanks to be filled out by the students with adequate linking words. After that, students received another feedback sheet with the filled-out rubric and the full exemplar and were instructed to use it for the final revision.</p>
16	Revision	<p>Did learners revise their work after using the rubric?</p> <p><input type="checkbox"/> No <input checked="" type="checkbox"/> Yes</p>
17	Extent of involvement	<p>How were learners involved in the rubric design and implementation?</p> <p><input checked="" type="checkbox"/> Students just received and used the rubric <input type="checkbox"/> Students were allowed to make small changes to the rubrics <input type="checkbox"/> Students made substantial changes <input type="checkbox"/> Students co-created the rubric <input type="checkbox"/> Other: Click here to add text</p>
18	Use of other instruments	<p>Were any additional instruments employed to further strengthen the intervention effects, or to make comparisons with the rubric? If so, please, explain the characteristics of those instruments</p> <p>We implemented another experimental group (EG2) who received in-text comments. We addressed structure and coherence in the left essay margin, while the right margin was reserved for selective surface-level error correction (spelling, word choice, grammar, tenses). Comment criteria aligned with rubric criteria (in EG1 and EG3) to maintain consistency. We prepared a pool of 70 comments for anticipated issues and six cues for additional scaffolds ahead of the intervention. We added four comments to this pool during the intervention, when our prepared comments did not cover an issue. We developed a decision-making flow chart to further systematize the feedback procedure and avoid overwhelming students with excessive comments, prioritizing crucial, higher-order aspects such as thesis statements and counter-arguments. Only when logically applicable did we comment on subsequent criteria, such as supporting details or rebuttals, avoiding redundancy and limiting comments.</p> <p>In a third experimental group (EG3), students received a rubric + exemplar and additional in-text comments in the margins of their essays. In doing so, we left it up to students to decide which feedback they prioritized in the revision or in what order and intensity they engaged with which form of feedback.</p>
19	Technology	<p>Was any type of technology used for the design and/or the implementation of the rubric? If so, please provide the details</p> <p>Click here to add text</p>
Outcomes		
20	Study Outcomes	<p>These variables are directly measured as outcomes of the rubric activity. Select all the options that apply to your study from the right column.</p> <p><input checked="" type="checkbox"/> Beliefs & perceptions: including perceptions of learning capacity to use the rubric (e.g., fairness, usefulness), metacognition and self-regulation, attitudes and beliefs (e.g., self-efficacy), teachers' perceptions/conceptions.</p>

		<input checked="" type="checkbox"/> Emotions and motivation: emotions experienced by learners (e.g., achievement emotions, social emotions, etc.) & motivational beliefs (e.g., learning motivation). <input checked="" type="checkbox"/> Performance: academic/domain specific performance, achievement, improved draft/work (i.e., revision). <input type="checkbox"/> Skills: quality of contribution to the group, professional behaviour, problem solving skills, work habits, interpersonal skills, metacognitive & self-regulatory skills. <input type="checkbox"/> Reliability of rubric: consistency of rubric scores among different raters (e.g., several teachers). <input type="checkbox"/> Validity of rubric: aspects related to testing the validity, such as content validity, comparing students and teachers' assessment, etc. <input type="checkbox"/> Other: Click here to add text
Moderators/mediators		
21 Moderators/mediators	Variables that are not usually manipulated but are taken into account when investigating rubrics. Select the variables that have been explored in your study from the right column.	<input checked="" type="checkbox"/> Gender: of the learner. <input checked="" type="checkbox"/> Academic characteristics: includes prior knowledge, prior performance, achievement level, GPA, finished high school, previous level of education, year of enrolment, etc. <input checked="" type="checkbox"/> Skills: self-regulated learning, self-efficacy, etc. <input checked="" type="checkbox"/> Age/grade level: of the learner. <input checked="" type="checkbox"/> Other: subtest on general cognitive abilities (figural reasoning), genre knowledge
Additional source information containing author names		

Appendix 2

Scales used to assess affective-motivational variables and feedback perceptions.

Scale (number of items)	Example item	Cronbach 's alpha			
		measurement point 1 during intervention	measurement point 2 during intervention	T1	T2
Writing self-efficacy (9)	I can formulate a thesis statement			0.911	0.904
Writing enjoyment (4)	I enjoy writing in English			0.912	0.945
Perceived quality of feedback (7)	The feedback has brought me closer to the learning goal	0.893	0.880		–
Perceived motivational impact of feedback (3)	The feedback made me enjoy revising my essay more	0.877	0.916		–
Summative evaluation of the feedback (4)	I would like to receive such feedback more often	–	–		0.906

Appendix 3: Example of rubric + exemplar used to provide feedback to EG1 and EG3 for writing task 2

Feedback for writing task 2 - Students who want to go abroad should attend a preparatory class

Learning goals:

- You can write well-structured, coherent argumentative essays.
- You can revise your essays using the feedback.

How to use the feedback sheet:

1. Read the feedback sheet.
2. Identify where the boxes **partly** or **not yet** are checked. These boxes show you where your essay needs revision. The model text on the right shows you what a good essay can look like.
3. Use the feedback to write an improved version of your essay.

Criteria for argumentative essays		yes	partly	not yet
Introduction	There is an introductory sentence (or several), [INTRO] which leads to the essay's topic.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	There is a thesis statement, [TH] which expresses your opinion about the topic.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	There is a transitional sentence, [TR] which says what will follow in the main body.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Main body	There is a counterargument, [CON] in opposition to the pro-arguments.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	There is a rebuttal, [BUT] which challenges the counterargument.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	There are at least two pro-arguments, [PRO] which support your thesis statement.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conclusion	There is support for each pro-argument, [S] such as an example, an explanation, an expert's opinion, or facts.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	There is a summary of the main arguments, [SUM] which briefly recaps your ideas.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	There is a rephrasing of the thesis statement, [RE] which uses different words than in the introduction.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Structural elements	There is a closing sentence, [CLOSE] such as a call to action, a piece of advice, a prediction, or a compromise.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	There are paragraph breaks after the introduction, when a new idea starts, and after the main body.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Language	There are linking words or phrases, which meaningfully connect your ideas.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Please check: <input type="checkbox"/> spelling, especially: _____ <input type="checkbox"/> grammar, especially: _____ <input type="checkbox"/> tense, especially: _____ <input type="checkbox"/> word choice, especially: _____			

Model text: Preparatory classes for exchange students are a good idea

During a student exchange, you encounter many new situations, which you might want to prepare for. This essay discusses the topic of preparatory classes for exchange students [INTRO]. Although there might be some disadvantages, in my opinion, all exchange students should attend such a class [TH]. There are many good reasons why, which I will consider in the following [TR].

Firstly, attending a preparatory class is a great opportunity to connect with other exchange students, who you would not meet otherwise [PRO]. This means that you can share your thoughts and fears about the exchange experience with people who are in the same situation as you, which makes going abroad less scary [S].

Although some people might say that you don't need preparatory classes because you can look up everything on the internet [CON], there are certain things that can be hard to organize by yourself, such as a visa [BUT].

Lastly, but most importantly, in a preparatory class you can learn strategies for dealing with unexpected situations [PRO]. For example, you learn how to put yourself in someone else's shoes, which may help you avoid hurting other people's feelings while you are abroad [S].

All in all, such a preparatory class has multiple benefits not only for the students, but for everyone else involved [SUM]. Therefore, every exchange student should take a preparatory class [RE]. It will help you make the most of your student exchange [CLOSE].

Toolbox
This toolbox sheet might be especially helpful for you:

- Tenses
- Elements of argumentative essays
- Vocabulary sheets
- Train of thought
- What is an argumentative essay?
- Linking words and phrases

Appendix 4

Pearson correlations between all variables among at-risk writers in EG 1–3 and CG1 ($n = 101$).

	1	2	3	4	5	6	7	8	9	10	11	12	
1	Writing quality English (T1)	1											
2	Deep-level features (Writing 1)	0.317***	1										
3	Deep-level features (Revision 1)	0.342***	0.548***	1									
4	Deep-level features (Writing 2)	0.167	0.509***	0.523***	1								
5	Deep-level features (Revision 2)	0.194	0.412***	0.535***	0.634**	1							
6	Writing self-efficacy (T1)	0.473***	0.213*	0.219*	0.068	0.162	1						
7	Writing self-efficacy (T2)	0.387***	0.244*	0.218*	0.216*	0.211*	0.622***	1					
8	Writing enjoyment (T1)	0.343***	0.038	0.054	0.108	0.076	0.560***	0.414***	1				
9	Writing enjoyment (T2)	0.249**	0.096	0.209*	0.151	0.212*	0.531***	0.612**	0.718***	1			
10	Perceived quality of feedback (during intervention)	0.233	0.103	0.180	0.125	0.119	0.262*	0.436***	0.183	0.389***	1		
11	Perceived motivational impact of feedback (during intervention)	0.314**	0.144	0.069	0.193	0.062	0.365**	0.547***	0.305**	0.508***	0.676***	1	
12	Summative evaluation of feedback (T2)	0.241**	0.112	0.196	0.337**	0.211	0.139	0.342**	0.142	0.304**	0.672***	0.597***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (two-tailed)

Appendix 5

Pearson correlations between all variables among developing writers in EG 1–3 and CG1 (*n* = 101).

	1	2	3	4	5	6	7	8	9	10	11	12
1 Writing quality English (T1)	1											
2 Deep-level features (Writing 1)	0.433***	1										
3 Deep-level features (Revision 1)	0.297**	0.534***	1									
4 Deep-level features (Writing 2)	0.291**	0.502***	0.476***	1								
5 Deep-level features (Revision 2)	0.030	0.392***	0.449***	0.735***	1							
6 Writing self-efficacy (T1)	0.020	0.260**	0.190	0.092	0.173	1						
7 Writing self-efficacy (T2)	0.199	0.365***	0.389***	0.312**	0.315**	0.681***	1					
8 Writing enjoyment (T1)	0.051	0.208*	0.183	0.109	0.125	0.583***	0.454***	1				
9 Writing enjoyment (T2)	-0.059	0.099	0.162	0.155	0.252*	0.476***	0.499***	0.683***	1			
10 Perceived quality of feedback (during intervention)	0.175	0.331**	0.381***	0.500***	0.381***	0.372**	0.509***	0.309**	0.362**	1		
11 Perceived motivational impact of feedback (during intervention)	0.047	0.152	0.184	0.078	0.115	0.350**	0.476***	0.461***	0.653***	0.601***	1	
12 Summative evaluation of feedback (T2)	0.150	0.327**	0.434***	0.552***	0.548***	0.249*	0.522***	0.359**	0.465***	0.723***	0.562***	1

p* < 0.05, *p* < 0.01, ****p* < 0.001 (two-tailed)

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors have no conflict of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abu-Rabia, S. (2003). The influence of working memory on reading and creative writing processes in a second language. *Educational Psychology, 23*(2), 209–222. <https://doi.org/10.1080/01443410303227>
- Anastasiou, D., & Michail, D. (2013). Exploring discordance between self-efficacy and writing performance among low-literate adult students. *Learning Disabilities: A Contemporary Journal, 11*(1), 53–87.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Erlbaum.
- Berry, D. N., Taylor, M. L., Nicolai, K. D., Koenka, A. C., Barnes, N., de Oliveira, L. C., & Zumbrunn, S. (2025). Reading and writing feedback perceptions among multilingual and monolingual learners across K-20 contexts: A systematic review. *Learning and Instruction, 95*, 102022. <https://doi.org/10.1016/j.learninstruc.2024.102022>
- Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. *ETS Research Report Series, 2011*(1), i–99. <https://doi.org/10.1002/j.2333-8504.2011.tb02241.x>
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education, 3*(22), 1–12. <https://doi.org/10.3389/educ.2018.00022>
- Brooks, C., Huang, Y., Hattie, J., Carroll, A., & Burton, R. (2019). What is my next step? School students' perceptions of feedback. *Frontiers in Education, 4*, 1–14. <https://doi.org/10.3389/educ.2019.00096>
- Bruning, R. H., & Kauffman, D. F. (2016). Self-efficacy beliefs and motivation in writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (eds.), *Handbook of writing research* 2nd ed., pp. 160–173. The Guilford Press, UK.
- Bruning, R. H., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology, 105*(1), 25–38. <https://doi.org/10.1037/a0029692>
- Busse, V. (2013). How do students of German perceive feedback practices at university? A motivational exploration. *Journal of Second Language Writing, 22*(4), 406–424. <https://doi.org/10.1016/j.jslw.2013.09.005>
- Busse, V., Krause, U. M., Parr, J. M., & Schubert, P. (2020). Developing secondary students' writing skills: Affective and motivational effects of a feedback intervention with learners of English as a foreign language. In F. Lenz, M. Frobenius, & R. Klattenberg (Eds.), *Classroom observation: Research interaction in English language teaching*, (pp. 245–265). Peter Lang.
- Busse, V., Graham, S., Müller, N., & Utesch, T. (2023). Understanding the interplay between text quality, writing self-efficacy and writing anxiety in learners with and without migration background. *Frontiers in Psychology, 14*, 1–14. <https://doi.org/10.3389/fpsyg.2023.1130149>

- Busse, V., Graham, S., & van Keer, H. (2024). Feedback to diverse language learners: Where are we and where do we need to go in literacy research? An introduction to the special issue on literacy and feedback. *Learning and Instruction*. Advance online publication. <https://doi.org/10.1016/j.learninstruc.2024.102035>
- Camacho, A., Alves, R. A., & Boscolo, P. (2021). Writing motivation in school: A systematic review of empirical research in the early twenty-first century. *Educational Psychology Review*, 33(1), 213–247. <https://doi.org/10.1007/s10648-020-09530-4>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Cheng, Y.-S. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing*, 13(4), 313–335. <https://doi.org/10.1016/j.jslw.2004.07.001>
- Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication*, 20(1), 99–118. <https://doi.org/10.1177/0741088303253572>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Council of Europe. Education Department. (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors*. Strasbourg: Council of Europe Publishing.
- Duijnhouwer, H., Prins, F. J., & Stokking, K. M. (2010). Progress feedback effects on students' writing mastery goal, self-efficacy beliefs, and performance. *Educational Research and Evaluation*, 16(1), 53–74. <https://doi.org/10.1080/13803611003711393>
- Ekholm, E., Zumbrunn, S., & DeBusk-Lane, M. (2018). Clarifying an elusive construct: A systematic review of writing attitudes. *Educational Psychology Review*, 30(3), 827–856. <https://doi.org/10.1007/s10648-017-9423-5>
- Ferris, D. (2011). *Treatment of error in second language student writing* (2nd edn.). Michigan series on teaching multilingual writers. University of Michigan Press.
- Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research*, 57(4), 481. <https://doi.org/10.2307/1170433>
- Fleckenstein, J., Jansen, T., Meyer, J., Trüb, R., Raubach, E. E., & Keller, S. (2024). How am I going? Behavioral engagement mediates the effect of individual feedback on writing performance. *Learning and Instruction*, 93, 1–10. <https://doi.org/10.1016/j.learninstruc.2024.101977>
- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication*, 37(1), 16–55. <https://doi.org/10.2307/357381>
- Galbraith, D. (2009). Cognitive models of writing. *German as a Foreign Language*, 2–3, 7–22.
- Graham, S., & Perin, D. (2007). Writing next: Effective strategies to improve writing of adolescents in middle and high schools: A report to Carnegie Corporation of New York. Alliance for Excellent Education. <https://education.illinoisstate.edu/downloads/casei/5-15-writingnext.pdf>
- Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational Psychologist*, 53(4), 258–279. <https://doi.org/10.1080/00461520.2018.1481406>
- Goossens, M., & Maeyer, S. (2018). How to obtain efficient high reliabilities in assessing texts: Rubrics vs comparative judgement. In E. Ras & A. E. Guerrero Roldán (eds.), *Technology enhanced assessment* (pp. 13–25). Springer International Publishing, Cham.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1–27). Erlbaum.
- Hayes, J. R., & Flower, L. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3–30). Lawrence Erlbaum.
- He, J., Liu, Y., Ran, T., & Zhang, D. (2023). How students' perception of feedback influences self-regulated learning: The mediating role of self-efficacy and goal orientation. *European Journal of Psychology of Education*, 38(4), 1551–1569. <https://doi.org/10.1007/s10212-022-00654-5>
- Heller, K. A., & Perleth, C. (2000). KFT 4-12+R: Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision [KFT 4-12+R: Cognitive ability test for grades 4-12, revision]. Beltz Test.

- Hetthong, R., & Teo, A. (2013). Does writing self-efficacy correlate with and predict writing performance? *International Journal of Applied Linguistics & English Literature*, 2(1), 157–167. <https://doi.org/10.7575/ijalel.v2n.1p.157>
- Hidi, S., Berndorff, D., & Ainley, M. (2002). Children's argument writing, interest and self-efficacy: An intervention study. *Learning and Instruction*, 12(4), 429–446. [https://doi.org/10.1016/S0959-4752\(01\)00009-3](https://doi.org/10.1016/S0959-4752(01)00009-3)
- Hinkel, E. (2011). *Handbook of research in second language teaching and learning* (Vol. 2). London: Routledge.
- Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing*, 10(3), 185–212. [https://doi.org/10.1016/S1060-3743\(01\)00038-8](https://doi.org/10.1016/S1060-3743(01)00038-8)
- Hyland, K. (1990). A genre description of the argumentative essay. *RELC Journal*, 21(1), 66–78. <https://doi.org/10.1177/003368829002100105>
- Hyland, K. (2003). *Second language writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511667251>
- Johnson, M. D. (2020). Planning in L1 and L2 writing: Working memory, process, and product. *Language Teaching*, 53(4), 433–445. <https://doi.org/10.1017/S0261444820000191>
- Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 355–365. <https://doi.org/10.1037/0278-7393.14.2.355>
- Khajavy, G. H., MacIntyre, P. D., & Barabadi, E. (2018). Role of the emotions and classroom environment in willingness to communicate. *Studies in Second Language Acquisition*, 40(3), 605–624. <https://doi.org/10.1017/S0272263117000304>
- Klassen, R. (2002). Writing in early adolescence: A review of the role of self-efficacy beliefs. *Educational Psychology Review*, 14, 173–203.
- Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, 21(4), 390–403. <https://doi.org/10.1016/j.jslw.2012.09.003>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.
- Lira Lorca, A., Peltzer, K., Busse, V., & Krause, U. M. (2024). Teaching argumentative writing in English as a foreign language: Insights into a learning unit including formative feedback. *JoSch - Journal Für Schreibwissenschaft*, 15(1), 7–24. <https://doi.org/10.3278/JOS2402W002>
- Lee, J. S. (2022). The role of grit and classroom enjoyment in EFL learners' willingness to communicate. *Journal of Multilingual and Multicultural Development*, 43(5), 452–468. <https://doi.org/10.1080/01434632.2020.1746319>
- Li, C., Wei, L., & Lu, X. (2023). Contributions of foreign language writing emotions to writing achievement. *System*, 116, 1–15. <https://doi.org/10.1016/j.system.2023.103074>
- Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science*, 42(4), 539–559. <https://doi.org/10.1007/s11251-013-9299-9>
- Lipnevich, A. A., Panadero, E., & Calistro, T. (2023). Unraveling the effects of rubrics and exemplars on student writing performance. *Journal of Experimental Psychology: Applied*, 29(1), 136–148. <https://doi.org/10.1037/xap0000434>
- McGarrell, H., & Verbeem, J. (2007). Motivating revision of drafts through formative feedback. *ELT Journal*, 61(3), 228–236. <https://doi.org/10.1093/elt/ccm030>
- Mills, M. (2014). Self-efficacy in second language acquisition. In S. Mercer & M. Williams (eds.), *Second language acquisition. Vol. 73. Multiple perspectives on the self in SLA* (pp. 6–22). Multilingual Matters.
- Müller, N., Utesch, T., & Busse, V. (2023). Qualität statt Quantität? Zum Zusammenhang von Schreibförderungs- und Feedbackpraktiken mit Textqualität unter Berücksichtigung von migrationsbedingter Mehrsprachigkeit [Quality over quantity? Linking writing support and feedback practices with text quality in consideration of migration-related multilingualism]. *Unterrichtswissenschaft*, 51(2), 169–198. <https://doi.org/10.1007/s42010-023-00173-2>
- Needels, M. C., & Knapp, M. S. (1994). Teaching writing to children who are underserved. *Journal of Educational Psychology*, 86(3), 339–349. <https://doi.org/10.1037/0022-0663.86.3.339>
- Newell, G. E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly*, 46(3), 273–304. <https://doi.org/10.1598/RRQ.46.3.4>

- Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, 46(5), 756–778. <https://doi.org/10.1080/02602938.2020.1823314>
- Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly*, 19(2), 139–158. <https://doi.org/10.1080/10573560308222>
- Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. *Educational Psychology Review*, 35(4), 1–38. <https://doi.org/10.1007/s10648-023-09823-4>
- Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, 35, 1–22. <https://doi.org/10.1016/j.edurev.2021.100416>
- Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing*, 15(2), 68–85. <https://doi.org/10.1016/j.asw.2010.05.004>
- Peltzer, K., Lira Lorca, A., Krause, U. M., Graham, S., & Busse, V. (2025). Effects of feedback on deep-level features of argumentative writing over multiple drafts: Insights from an intervention study with secondary EFL students. *Reading and Writing*. Advance online publication. <https://doi.org/10.1007/s11145-025-10638-8>
- Peltzer, K., Lira Lorca, A., Krause, U. M., & Busse, V. (2024). Effects of formative feedback on argumentative writing in English and cross-linguistic transfer to German. *Learning and Instruction*, 92, 1–16. <https://doi.org/10.1016/j.learninstruc.2024.101935>
- Peltzer, K., Siekmann, L., Parr, J. M., & Busse, V. (2022). What beliefs about writing guide EFL curricula? An analysis of relevant policy documents for teaching English at German secondary schools. *Zeitschrift Für Erziehungswissenschaft*, 25(6), 1363–1387. <https://doi.org/10.1007/s11618-022-01089-w>
- Scherer, S., Graham, S., & Busse, V. (2024). How effective is feedback for L1, L2, and FL learners' writing? A meta-analysis. *Learning and Instruction*, 93, 1–17. <https://doi.org/10.1016/j.learninstruc.2024.101961>
- Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18(3), 337–354. <https://doi.org/10.1006/ceps.1993.1024>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Siekmann, L., & Busse, V. (2022). Process-oriented writing and formative feedback in EFL classes: A comparison of teachers' and learners' perceptions. In E. Wilden, L. Alfes, K. F. Cantone, S. S. Çikrikçi, & D. Reimann (Eds.), *Beiträge zur Fremdsprachenforschung*, vol. 17. *Standortbestimmungen in der Fremdsprachenforschung [Contributions to foreign language research, Vol. 17, Determining the status quo in foreign language research]* (pp. 274–290). Schneider wbv.
- Siekmann, L., Parr, J. M., & Busse, V. (2022). Structure and coherence as challenges in composition: A study of assessing less proficient EFL writers' text quality. *Assessing Writing*, 54, 1–13. <https://doi.org/10.1016/j.asw.2022.100672>
- Siekmann, L., Parr, J. M., van Ophuysen, S., & Busse, V. (2023). Text quality and changing perceptions of teacher feedback and affective-motivational variables: A study with secondary EFL students. *Frontiers in Education*, 8, 1–16. <https://doi.org/10.3389/educ.2023.1171914>
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 Writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657–677. <https://doi.org/10.2307/3587400>
- Stevenson, M., Schoonen, R., & de Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*, 15(3), 201–233. <https://doi.org/10.1016/j.jslw.2006.06.002>
- Sun, T., Wang, C., Lambert, R. G., & Liu, L. (2021). Relationship between second language English writing self-efficacy and achievement: A meta-regression analysis. *Journal of Second Language Writing*, 53, 1–14. <https://doi.org/10.1016/j.jslw.2021.100817>
- Tahmouresi, S., & Papi, M. (2021). Future selves, enjoyment and anxiety as predictors of L2 writing achievement. *Journal of Second Language Writing*, 53, 100837. <https://doi.org/10.1016/j.jslw.2021.100837>
- Teng, L. S., Sun, P. P., & Xu, L. (2018). Conceptualizing writing self-efficacy in English as a foreign language contexts: Scale validation through structural equation modeling. *TESOL Quarterly*, 52(4), 911–942. <https://doi.org/10.1002/tesq.432>

- Teng, M. F., & Wang, C. (2023). Assessing academic writing self-efficacy belief and writing performance in a foreign language context. *Foreign Language Annals*, 56(1), 144–169. <https://doi.org/10.1111/flan.12638>
- To, J., Panadero, E., & Carless, D. (2022). A systematic review of the educational uses and effects of exemplars. *Assessment & Evaluation in Higher Education*, 47(8), 1167–1182. <https://doi.org/10.1080/02602938.2021.2011134>
- Toulmin, S. E. (2003). *The uses of argument (2nd edition)*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511840005>
- Trace, J., Meier, V., & Janssen, G. (2016). “I can see that”: Developing shared rubric category interpretations through score negotiation. *Assessing Writing*, 30, 32–43. <https://doi.org/10.1016/j.asw.2016.08.001>
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, 78(4), 751–796. <https://doi.org/10.3102/0034654308321456>
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners’ agentic engagement with feedback: A systematic review and a taxonomy of reciprocity processes. *Educational Psychologist*, 52(1), 17–37. <https://doi.org/10.1080/00461520.2016.1207538>
- Woodrow, L. (2011). College English writing affect: Self-efficacy and anxiety. *System*, 39(4), 510–522. <https://doi.org/10.1016/j.system.2011.10.017>
- Zarrinabadi, N., & Rezazadeh, M. (2020). Why only feedback? Including feed up and feed forward improves non-linguistic aspects of L2 writing. *Language Teaching Research*, 27(3), 575–592. <https://doi.org/10.1177/1362168820960725>
- Zhao, P., Zhu, X., Yao, Y., & Liao, X. (2023). Ideal L2 self, enjoyment, and strategy use in L2 integrated writing: A self-regulatory learning perspective. *System*, 115, 1–10. <https://doi.org/10.1016/j.system.2023.103033>
- Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22(1), 73–101. <https://doi.org/10.1006/ceps.1997.0919>
- Zumbrunn, S., Marrs, S., Broda, M., Ekholm, E., DeBusk-Lane, M., & Jackson, L. (2019). Toward a more complete understanding of writing enjoyment: A mixed methods study of elementary students. *AERA Open*, 5(2), 1–16. <https://doi.org/10.1177/2332858419850792>
- Zumbrunn, S., Marrs, S., & Mewborn, C. (2016). Toward a better understanding of student perceptions of writing feedback: A mixed methods study. *Reading and Writing*, 29(2), 349–370. <https://doi.org/10.1007/s11145-015-9599-3>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Katrin Peltzer¹  · Alina Lira Lorca² · Ulrike-Marie Krause² · Steve Graham³ · Ernesto Panadero^{4,5}  · Vera Busse¹ 

✉ Katrin Peltzer
katrin.peltzer@uni-muenster.de

Alina Lira Lorca
alina.lira.lorca@uni-oldenburg.de

Ulrike-Marie Krause
ulrike.krause@uni-oldenburg.de

Steve Graham
steve.graham@asu.edu

Ernesto Panadero
ernesto.panadero@dcu.ie

Vera Busse
vbusse@uni-muenster.de

- ¹ Department of Education, University of Münster, Schlossplatz 2 Georgskommende 33, Room C207, 48149 Münster, Germany
- ² Department of Educational Sciences, Carl Von Ossietzky Universität Oldenburg, Ammerländer Heerstraße 114-118, 26129 Oldenburg, Germany
- ³ Arizona State University, PO Box 871811, Tempe, AZ 85287, USA
- ⁴ Centre for Assessment Research, Policy and Professional Practice in Education (CARPE), Dublin City University, DCU St. Patrick's Campus, Dublin 9, Ireland
- ⁵ Education Regulated Learning and Assessment (ERLA) Group, Faculty of Education and Sport, Deusto University, Bilbao, Spain