

## Article

# Promoting Reflection Skills of Pre-Service Teachers—The Power of AI-Generated Feedback

Florian Hofmann \*, Tina-Myrica Daunicht, Lea Plöchl  and Michaela Gläser-Zikuda 

Research and Teaching Unit for School Education and Instructional Research, Institute for Educational Science, Faculty of Humanities, Social Sciences, and Theology, Friedrich-Alexander-Universität Erlangen-Nürnberg, 90478 Nuremberg, Germany; tina-myrica.daunicht@fau.de (T.-M.D.); lea.ploessl@fau.de (L.P.); michaela.glaeser-zikuda@fau.de (M.G.-Z.)

\* Correspondence: florian.hofmann@fau.de

## Abstract

Reflection skills are a key but challenging element in teacher training. Feedback on reflective writing assignments can improve reflection skills, but it is affected by challenges (high variability in judgments and time investment). AI-generated feedback offers many options. Therefore, the aim of this study was to examine the potential of AI-generated feedback compared to that provided by lecturers for developing reflective skills. A total of 93 randomly selected pre-service teachers (70% female) in a course at a German university wrote two reflections and received feedback from either lecturers or ChatGPT 4.0 based on the same prompts. Pre-service teachers' written reflections were assessed, and an online questionnaire based on standard instruments was applied. Control variables included metacognitive learning strategies and reflection-related dispositions. Based on a linear mixed model, the main effects on reflective skills were identified for time ( $\hat{\beta} = 0.41$ ,  $p = 0.003$ ) and feedback condition ( $\hat{\beta} = -0.42$ ,  $p = 0.032$ ). Both forms of feedback similarly fostered reflective skills over time, with academic self-efficacy emerging as a pertinent disposition ( $\hat{\beta} = 0.25$ ,  $p = 0.014$ ). The limitations of this study and implications for teacher training are discussed.



Academic Editors: Kira Elena Weber and Hendrik Lohse-Bossenz

Received: 1 July 2025

Revised: 20 August 2025

Accepted: 12 September 2025

Published: 3 October 2025

**Citation:** Hofmann, F., Daunicht, T.-M., Plöchl, L., & Gläser-Zikuda, M. (2025). Promoting Reflection Skills of Pre-Service Teachers—The Power of AI-Generated Feedback. *Education Sciences*, 15(10), 1315. <https://doi.org/10.3390/educsci15101315>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** reflection skills; artificial intelligence (AI); feedback

## 1. Introduction

Teachers' professional development is becoming increasingly important. Teachers need not only to ensure that students' academic achievement is improving but also to develop learning environments in schools based on 21st-century skills (Benade, 2017; Chan & Lee, 2021; Idel et al., 2020). The ability to (self-)reflect is seen as a core concept for almost all professional trainings (Chan & Lee, 2021; Ryan, 2013). For example, reflection processes have been shown to positively influence decision-making and problem-solving strategies (Sudirman et al., 2024). In particular, in professional teacher training, the ability to reflect is a core skill (Borko et al., 1997; Albert, 2016). Reflection serves as an important bridge between theoretical knowledge and professional action (Korthagen, 2018). Teachers must take an active role in their ongoing professional development process and be able to analyze and reflect on their skills (Albert, 2016). The ability to reflect is also a core component of self-directed (Zimmerman, 2000) and lifelong learning (Abdalina et al., 2022). (Self-)reflection processes are especially relevant for developing metacognitive skills (Desautel, 2009). However, research has shown that reflection skills are often poorly developed among

(pre-service) teachers (Zhang et al., 2024). Therefore, teacher training programs need approaches that systematically support pre-service teachers' reflection skills (Artmann & Herzmann, 2016). Feedback is particularly vital for learning as well as for professional development in general (Darling-Hammond & Bransford, 2007; Korthagen & Kessels, 1999). Implementing feedback in teacher training is associated with two major challenges for lecturers: first, they often perceive the evaluation of pre-service teachers' reflections as extremely complex (Ryan, 2013), and second, providing individual feedback requires considerable resources (Fleckenstein et al., 2023; Ullmann, 2015). As a result, approaches that incorporate computer-assisted and automated feedback into teaching and learning processes are increasingly being developed and tested (Chong et al., 2020; Fleckenstein et al., 2023; Gibson et al., 2016; Wongvorachan et al., 2022; Wulff et al., 2023). Artificial intelligence (AI) is creating new opportunities, particularly in this context (Zhang et al., 2024), but it also introduces new challenges and problem areas, for example, regarding privacy and data protection (Sebastian, 2023) or AI ethics (Huang, 2023). Given the increasing importance of AI-supported feedback concepts in teaching and learning, this study compared AI-supported feedback with feedback provided by lecturers and examined their effectiveness in promoting pre-service teachers' reflection skills.

## 2. Theoretical Background

### 2.1. Reflection and Reflective Skills in Teacher Education

The ability to reflect has been deemed essential for tackling the growing challenges in schools (Chan & Lee, 2021; Ryan, 2013). In Germany, the KMK Standards for Teacher Education (KMK—Ständige Konferenz der Kultusminister der Bundesrepublik Deutschland, 2022) outline quality assurance guidelines that specifically promote the cumulative development of experience and skills through theory-based reflection. Reflection processes are primarily intended to support the crucial link between theory and practice (Korthagen, 2018) and to constructively integrate the various professional domains of teaching (Albert, 2016). Above all, reflection should contribute to the primary goal of initiating change and innovation: fostering a reorientation towards improved and individualized learning through new insights (Saric & Steh, 2017). Reflection, particularly in the context of teacher training, can be theoretically situated in various ways (Korthagen, 1999). Many academic discussions build upon the ideas of Dewey (1910/2022), which are meaningfully supplemented by the “reflective practitioner” approaches of Argyris and Schön (2024). The definition by Lenske and Lohse-Bossenz (2023) also draws on these principles, defining professional reflection in pedagogical contexts as “an occasion-related mental process that, with self-reference, aims at an expanded understanding of pedagogical practice” (p. 118). Given the many different classifications and forms of implementing reflection (Aeppli & Lötscher, 2016), teacher training generally focuses on analyzing situations or events (Korthagen, 1999) in order to generate options for future action (Schön, 1987). Importantly, it addresses not only cognitive reflection processes and options for action but also motivational and volitional aspects (Leonhard & Rihm, 2011). Almost all models depicting reflection processes build on these fundamental assumptions and differentiate the complex process into distinct phases and sub-competencies (e.g., Aeppli & Lötscher, 2016; Von Aufschnaiter et al., 2019; Blömeke et al., 2015). Given the heterogeneous nature of the teaching profession (e.g., different school types, subjects, grade levels, degrees, etc.), the model by Von Aufschnaiter et al. (2019) provides a particularly suitable framework because it is designed to be domain-independent and dynamically accommodates varying emphases and focuses depending on the context. This model divides the reflection process into (1) “reflection-related dispositions” (fundamental characteristics underlying situation-related thought processes); (2) “reflection-related thought processes” (a situation that serves as a trigger for reflection,

ideally within a prepared learning environment); and (3) “reflection-related performance” (the quality or output of the reflection process) (Von Aufschnaiter et al., 2019, p. 152). According to this model, in addition to a specific and well-explained impulse for reflection, a successful reflection process primarily depends on reflection-related dispositions. These include, for example, prior experience with reflection, attitudes and beliefs about its relevance, and the individual’s self-efficacy, i.e., the confidence to carry out the reflective process independently (Von Aufschnaiter et al., 2019). Empirical results show that previous experience with reflection is related to reflection skills (Fraij & Kirschner, 2017), and that self-efficacy expectations influence reflective capability (Stender et al., 2021). The influence of attitudes on reflection and the necessity of developing reflection skills (Göbel & Neuber, 2020; Stender et al., 2021) is also well supported and aligns with findings from portfolio research and the importance of attitudes of pre-service teachers towards portfolio work (Feder et al., 2021).

Pre-service teachers often struggle with reflection and tend to achieve only low levels of reflective depth (Zhang et al., 2024). Consequently, several studies have investigated the conditions for successfully implementing reflection in teacher training (e.g., Hansen & Rachbauer, 2018; Hatton & Smith, 1995; Jaeger, 2013; Jones & Jones, 2013; Saric & Steh, 2017). It seems essential to systematically foster the development of reflective skills from the beginning of teacher training through targeted approaches (Artmann & Herzmann, 2016). Written engagement with the subject of reflection has proven to be an effective strategy for developing reflective skills (Häcker, 2019; Sudirman et al., 2024). Moreover, students require various forms of support to develop this complex skill; most frequently mentioned are detailed explanations, examples, feedback, and specific workshops (Ross et al., 2024). These measures are necessary to positively influence enthusiasm and attitudes toward reflective writing and to help students overcome difficulties and obstacles (Sudirman et al., 2024).

Different models can be used to assess written reflections (Boyd & Fales, 1983; Gibbs, 1988; Gore & Zeichner, 1991; Hatton & Smith, 1995). To determine levels of reflection, one usually attempts to reduce the complexity of reflections by employing depth or breadth models, or to classify the reflective thoughts according to specific criteria (Ullmann, 2019). Research on teacher education often draws on the depth model of Hatton and Smith (1995), which has been adapted multiple times (e.g., Riel, 2022; Zhang et al., 2023, 2024). In university-based teacher training, contextually tailored criteria grids specifying quality characteristics of successful reflections are commonly used for evaluation (Von Aufschnaiter et al., 2019; Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg, LI Hamburg, 2020). These criteria grids serve as the basis for assessment, feedback, or (self-)reflection questionnaires (e.g., Berger, 2022).

## 2.2. Feedback and AI

Numerous meta-analyses have shown positive effects of feedback on promoting learning processes and performance (e.g., Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Feedback is also highly relevant for reflective writing (Darling-Hammond & Bransford, 2007; Korthagen & Kessels, 1999), and its positive effects have already been demonstrated (Jahncke et al., 2018; Schellenbach-Zell et al., 2023). However, university lecturers consider evaluating written reflections extremely complex and time-consuming (Ryan, 2013). Furthermore, due to limited resources, it is often difficult to provide timely, detailed, and didactically prepared feedback on reflections to a larger number of students (Deeva et al., 2021; Fleckenstein et al., 2023; Ullmann, 2015).

Rapid technological advances, especially in artificial intelligence (AI), are fundamentally transforming the educational system (Zhai et al., 2021). Various computer-assisted and automated methods for feedback in learning and instruction have been developed (Chong

et al., 2020; Fleckenstein et al., 2023; Gibson et al., 2016; Suvendu & Deb, 2024; Wongvorachan et al., 2022; Wulff et al., 2023). AI approaches based on (open) large language models (LLMs), such as ChatGPT, are becoming increasingly popular (Chang et al., 2023). These models can be integrated into learning and instruction, for example, by providing feedback to students in instructional settings where feedback was previously not feasible due to the lack of this technological advancement (Deeva et al., 2021; Ullmann, 2019). The use of LLMs enables timely, personalized, and frequent feedback (Russell & Korthagen, 2013); however, lecturers need to know which systems can be applied in which contexts, and they must possess the skills to use them accordingly (Deeva et al., 2021). Initial trials mostly employ relatively simple task formats to test machine-generated feedback. Such automatic feedback systems have already proven effective in these areas, as (moderate) effects on the respective competence development have been observed (Fleckenstein et al., 2023).

However, reflections pose an enormous challenge for analysis and evaluation due to their complexity and multi-perspective nature (Körkkö et al., 2016; Poldner et al., 2014; Ullmann, 2019). The enormous substantial innovations in LLM training over the past two years have now enabled meaningful reflection assessment (Nehyba & Štefánik, 2023; Wulff et al., 2023). These advances allow the use of AI feedback on reflections in learning and instruction (Barthakur et al., 2022; Jung et al., 2022; Zhang et al., 2024), making it possible for a large number of students to receive feedback on their reflections promptly, or even during the work phase (Kovanović et al., 2018; Ullmann, 2019).

While there is great potential associated with the development of artificial intelligence (Trust et al., 2023; Winkel, 2025; Wongvorachan et al., 2022), including in the field of teacher training (Van den Berg & Du Plessis, 2023), it also brings new obstacles and challenges (Chang et al., 2023; Eden et al., 2024; Suvendu & Deb, 2024). Open AI products such as ChatGPT in particular often face the issue that data protection aspects are not yet fully addressed and privacy is insufficiently safeguarded (Huang, 2023; Sebastian, 2023). Thus, innovations and improvements in the context of AI ethics are urgently needed (Sebastian, 2023; Winkel, 2025). This also implies that teachers, learners, and the entire educational landscape must develop competencies in these areas if technological advances are to be used effectively without compromising privacy (Buckingham Shum et al., 2023; Sperling et al., 2024; Watini et al., 2024; Wiese et al., 2025). Furthermore, new, technology-supported concepts also require new implementation strategies (Eden et al., 2024; Suvendu & Deb, 2024) as well as new pedagogical and didactic approaches to designing learning environments (Bearman & Ajjaw, 2023; Chang et al., 2023; Deeva et al., 2021; Eden et al., 2024; Li & Kim, 2024; Van den Berg & Du Plessis, 2023; Wiese et al., 2025). In addition to the lecturers, students also need to adapt to make meaningful use of AI feedback: when well-designed learning environments are in place, the challenge lies less in the ability to use automatic feedback systems and more in the ability to integrate them meaningfully into (self-regulated) learning processes and to critically evaluate the results (Bearman & Ajjaw, 2023; Li & Kim, 2024). Furthermore, skills such as creativity, teamwork, and problem-solving, which cannot (currently) be performed by AI systems, must be increasingly emphasized and developed (George, 2023).

This study focuses on the significant possibilities offered by AI tools in teaching and learning processes (Eden et al., 2024; Trust et al., 2023; Van den Berg & Du Plessis, 2023). In particular, the option of generating automatic feedback (Shaik et al., 2022; Wongvorachan et al., 2022) promises considerable potential, especially in connection with the complex skill of reflection (e.g., Chang et al., 2023).

Nevertheless, the effects of AI-based feedback compared to instructor feedback on the development of learners' reflection skills have not yet been investigated.

### 3. Research Questions

Considering the current state of research outlined above, it can be considered certain that pre-service teachers should possess or develop reflective competence. However, it remains unclear how this competence can be most effectively imparted. A promising approach is feedback, as the fundamental effects of feedback on learners' competency development are well-established. Positive effects of feedback have also been repeatedly demonstrated in the specific area of reflective ability. The question of whether AI-generated feedback similarly enhances the quality of pre-service teachers' written reflections remains to be clarified, particularly in comparison to feedback provided by lecturers. Furthermore, we aim to investigate whether reflection-related dispositions influence the use of (AI-supported) feedback. The current study therefore addresses the following research questions:

RQ 1: What is the effect of AI-supported feedback compared to instructor feedback on the reflection level in pre-service teachers' written reflections?

RQ 2: How do reflection-related dispositions influence the reflection level in pre-service teachers' written reflections?

### 4. Method

#### 4.1. Participants

A total of 93 pre-service teachers from four different teacher training programs at a German university participated in this study. The instructional concept, developed to improve reflection skills, was integrated into a basic module in the field of educational science. In this module, participants encountered, for the first time, a learning concept involving interaction with AI. Data were collected at the beginning (t1) and midpoint (t2) of the summer semester of 2024. A total of 93 pre-service teachers participated in the initial survey at t1, distributed across teacher training programs as follows: primary school (44.1%), lower secondary school (7.5%), intermediate secondary school (23.7%), and academic secondary school (24.7%). Eight participants discontinued the course, resulting in a final sample of the study with eighty-five pre-service teachers at t2 (see Table 1).

**Table 1.** Sociodemographic sample characteristics.

| Characteristic           | Feedback <sub>Instructor</sub> |          | Feedback <sub>AI</sub> |          | Sample          |          |
|--------------------------|--------------------------------|----------|------------------------|----------|-----------------|----------|
|                          | <i>M (SD)/%</i>                | <i>n</i> | <i>M (SD)/%</i>        | <i>n</i> | <i>M (SD)/%</i> | <i>n</i> |
| Age                      | 21.82 (3.24)                   | 45       | 21.00 (2.21)           | 48       | 21.40 (2.27)    | 93       |
| Gender                   |                                |          |                        |          |                 |          |
| Female                   | 66.7%                          | 30       | 72.9%                  | 35       | 69.9%           | 65       |
| Male                     | 33.3%                          | 15       | 25.0%                  | 12       | 29.0%           | 27       |
| Other                    | -                              | 0        | 2.1%                   | 1        | 1.1%            | 1        |
| Teacher Training Program |                                |          |                        |          |                 |          |
| Primary                  | 46.7%                          | 21       | 41.7%                  | 20       | 44.1%           | 41       |
| Lower Secondary          | 6.7%                           | 3        | 8.3%                   | 4        | 7.5%            | 7        |
| Intermediate Secondary   | 17.8%                          | 8        | 29.2%                  | 14       | 23.7%           | 22       |
| Academic Secondary       | 28.8%                          | 13       | 20.8%                  | 10       | 24.7%           | 23       |

#### 4.2. Design

##### 4.2.1. Procedure

The current study is part of the research project PetraKIP ("Personal Transparent AI-Based Portfolio for Teacher Training") funded by the German Federal Ministry of Education and Research.

To promote the pre-service teachers' reflection skills, an instructional concept was developed as part of a basic teacher training module in the domain of school education.

The concept comprised a lecture, a seminar, and two individual learning phases covering two topics (1. Diagnostics and 2. Classroom Management). The individual learning phase entailed the subsequent processing of a case study, each presenting a problem situation from the school context. After the first survey (t1), the participants submitted a written reflection on each case study over the semester, for which they received feedback.

Following the first written reflection, 48 participants were randomly assigned to receive AI-generated feedback, while 45 participants received lecturer-provided feedback. The pre-service teachers could then use this feedback when preparing the second written reflection. After submitting the second reflection (t2), all participants received AI-generated feedback on their work. By comparing the two feedback groups from t1 to t2, it was possible to evaluate the effects of AI-supported feedback versus lecturer feedback (RQ 1), while also considering the influence of reflection-related dispositions (RQ 2). Table 1 presents sociodemographic information for both feedback groups as assessed at t1.

#### 4.2.2. Instruments

##### Feedback Grid and Reflection Scores

To assess the quality of pre-service-teachers' written reflections, a grid with quality criteria was used (e.g., [Von Aufschnaiter et al., 2019](#); [Philipp, 2023](#)). This grid served as the basis both for an evaluation form for instructors and as a prompt for ChatGPT 4.0 ([Berger, 2022](#)) (see Appendix A for the full prompt). The starting point for the evaluation criteria was the compilation of quality characteristics according to [Meyer-Siever and Levin \(2018\)](#), which identify four major categories: (1) analysis based on theoretical models and empirical findings, (2) consideration of different perspectives, (3) developing alternative courses of action or considerations for future action, and (4) relating to one's professionalization. An existing instrument built on this foundation and already successfully used is the form developed by the State Institute for Teacher Training and School Development Hamburg ([Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg, LI Hamburg, 2020](#)). Their framework originally included the following ten criteria: (1) Linguistic clarity, (2) Structure, (3) Description of the experience, (4) Theoretical relevance, (5) Professional accuracy, (6) Richness of perspectives, (7) Own viewpoints (and their justification), (8) Self-critical perspective, (9) Constructive solutions and conclusions, and (10) Systemic perspective (linking of the individual aspects and perspectives). This framework was slightly modified because some criteria, particularly those relevant for practical use, were not considered or required more specific explanation or supplementation. A comparison with other quality criteria lists (e.g., [Von Aufschnaiter et al., 2019](#); [Hansen & Rachbauer, 2018](#); [Leonhard & Rihm, 2011](#); [Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg, LI Hamburg, 2020](#); [Philipp, 2023](#)) informed these adjustments. For example, the highly detailed compilation by [Von Aufschnaiter et al. \(2019\)](#) provided the criterion "scope", while "structure of reflection" was differentiated further by including explicitly stated reflection goals (including their justification). Furthermore, some criteria were reformulated (e.g., "describing the experience" became "object of reflection") or combined due to a lack of clarity in practice (e.g., "own perspectives" and "self-critical perspectives").

Thus, for this study, the following criteria were included in the grid: (1) Linguistic design, (2) Structure, (3) Scope, (4) Formulation of goals, (5) Justification of these goals, (6) Reference to the object of reflection, (7) Use of professional knowledge, (8) Own perspectives, (9) Different temporal perspectives, (10) Richness of perspectives, (11) Linking of the individual aspects and perspectives, and (12) Further thoughts and conclusions. According to [Leonhard and Rihm \(2011\)](#), a suitable instrument should reflect both the structure of the construct of reflective ability and the corresponding levels of expression. Each listed criterion was assessed on a scale of 0–3 ([Philipp, 2023](#)), with a maximum score

of 36. Detailed coding guidelines for instructors were developed and iteratively refined to ensure uniform and consistent scoring.

Most coders had prior experience in grading and providing feedback on reflective writing from previous semesters of the project. To further ensure consistency, all coders participated in a three-hour re(training) session each semester, which was led by the main author and focused on the twelve criteria and distinction of quality levels (0–3). After training, three coders independently rated a subset of reflection texts, but they were encouraged to consult with one another via chat exchange in cases of uncertainty. A second team of three coders conducted an independent second round of coding with similar opportunities for collegial exchange.

Interrater reliability was then assessed based on 540 coded segments (12 criteria across 45 pre-service teachers' written reflections in the instructor feedback group at t1), each rated by 2 of  $k = 6$  university instructors. Following [Shrout and Fleiss \(1979\)](#), intraclass correlation coefficients were calculated in R using a one-way random effects model for single measures (psych: ICC, [Revelle, 2025](#)). The resulting interrater reliability was excellent (ICC (1,1) = 0.93, 95% CI [0.93, 0.94]) according to benchmarks by [Koo and Li \(2016\)](#), supporting the use of reflection scores in subsequent analyses.

#### Online Questionnaire

All additional data were collected as part of an online survey at the beginning of the semester (t1) and after completing the first reflective writing assignment (t2). Apart from socio-demographic data, the following reflection-related dispositions were collected:

##### (1) Metacognitive Learning Strategies (t1)

Metacognitive Learning Strategies (MCLS) encompass the dimensions of planning, monitoring, and regulation in the context of learning activities. The scale was adapted from the German LIST inventory (Learning Strategies in Higher Education, [Wild & Schiefele, 1994](#)). Responses were recorded on a 5-point Likert scale ranging from 1 (not at all true) to 5 (completely true). The scale comprises 18 items and demonstrated good internal consistency, with Cronbach's  $\alpha = 0.85$ , 95% confidence interval (CI) [0.81, 0.90],  $n = 93$ . A sample item is: "If I do not understand a text upon first reading, I go through it again step by step."

##### (2) Academic Self-Efficacy (t1)

Academic self-efficacy (ASE) was measured using a six-item Likert scale based on the work of [Schiefele et al. \(2002\)](#). Responses were rated on a 5-point scale ranging from 1 (not at all true) to 5 (completely true), in response to items such as: "Even if an exam is very difficult, I know what I have to do to pass it." The scale indicated good reliability, with Cronbach's  $\alpha = 0.89$ , 95% CI [0.86, 0.92],  $n = 93$ .

##### (3) Attitude towards reflective writing tasks (t1)

Positive attitude towards reflective writing tasks (ARW) was assessed using a self-developed 5-point Likert scale, with response options ranging from 1 (not at all true) to 5 (completely true). The brief scale consisted of three items (e.g., "I believe that engaging in written reflections during teacher training enhances reflection on personal development.") and yielded good reliability estimates (Cronbach's  $\alpha = 0.79$ , 95% CI [0.72, 0.87],  $n = 93$ ).

##### (4) Previous experience with reflective writing tasks (t1)

Pre-service teachers' prior experience with reflective writing tasks (ERW) was measured via a self-developed single item: "I have already engaged in written reflection tasks prior to attending the Module 1 school pedagogy course." Responses could be given on a scale from 1 (not at all true) to 5 (completely true).

### (5) Feedback Engagement (t2)

To assess pre-service teachers' engagement with the feedback provided on their first reflective writing assignment (FE, treatment check), we employed a self-developed three-item Likert scale (e.g., "I carefully considered the feedback on my first reflection before composing the second one."). Response options ranged from 1 (not at all true) to 5 (completely true). The scale demonstrated acceptable reliability (Cronbach's  $\alpha = 0.60$ , 95% CI [0.46, 0.74],  $n = 84$ ).

### 4.3. Data Analysis

In the first step, eight instructors assessed written reflections from both feedback groups (AI vs. instructor) and both measurement points (t1 and t2), following the coding guideline. These reflection scores (RSs) served as the basis for further analysis.

In the second step, a linear mixed effects model (LMM) with random intercepts was applied to account for individual variability and the dependency of repeated measures in the RS. LMMs allow for more precise estimation of fixed and random effects and offer increased statistical power, even with small sample sizes (Brauer & Curtin, 2018).

The dataset was transformed into long format (two rows per participant) to model differences between measurement occasions, with RS collapsed into a single variable and measurement time point added as a factor (Time; 0 = t1, 1 = t2). The RS variable was z-standardized and grand-mean centered for better comparability and interpretation of results. The final model included RS as an outcome predicted by time, feedback condition (including their interaction), and reflection-related dispositions (metacognitive learning strategies, academic self-efficacy, attitude towards reflective writing tasks, previous experience with reflective writing tasks, and engagement with provided feedback). Analyses were conducted in R (R Core Team, 2025) using the lme4 package with restricted maximum likelihood estimation (Bates et al., 2015).

## 5. Results

### 5.1. Descriptive Statistics

Table 2 presents the means (M) and standard deviations (SD) for the first (t1) and second (t2) measurement points across both feedback groups. For the first measurement point, the table lists descriptive values for reflection scores (RS1), metacognitive learning strategies (MCLS), and reflection-related disposition variables (academic self-efficacy, ASE; experience with reflective writing, ERW; attitude towards reflective writing, ARW). For the second measurement point, feedback engagement (FE) and reflection scores (RS2) are reported. Both groups scored very high on academic self-efficacy (ASE) and moderately high on metacognitive learning strategies (MCLS) and (positive) attitudes towards reflective writing (ARW). Most students had little prior experience with reflective writing (ERW), although the standard deviation indicates considerable variability within the sample. Descriptive statistics further indicate that metacognitive learning strategies (MCLS) and reflection-related disposition variables (ASE, ERW, ARW) are similar across the two groups, with the main differences observed in reflection scores (RS1 and RS2).

Reflection scores were moderately correlated between measurement points,  $r = 0.56$ ,  $p < 0.001$ . Among the predictor variables considered for inclusion in the LMM, the most significant intercorrelation coefficients fell into the small effect range ( $0.44 > r > 0.23$ ,  $p < 0.05$ ), except for metacognitive learning strategies (MCLS; see also Table 3). MCLS showed a strong correlation with academic self-efficacy (ASE) at t1,  $r = 0.64$ ,  $p < 0.001$ ; a moderate correlation with feedback engagement (FE) at t2,  $r = 0.43$ ,  $p < 0.001$ ; and a weak correlation with attitude towards reflective writing (ARW) at t1,  $r = 0.27$ ,  $p = 0.009$ . Correlat-

tions between MCLS and reflection scores at both measurement points were non-significant. Due to the risk of multicollinearity, MCLS was excluded as a predictor in the LMM.

**Table 2.** Means and standard deviations for all variables across both measurement points.

| Variable                     | Feedback <sub>Instructor</sub> |          | Feedback <sub>AI</sub> |          | Sample        |          |
|------------------------------|--------------------------------|----------|------------------------|----------|---------------|----------|
|                              | <i>M (SD)</i>                  | <i>n</i> | <i>M (SD)</i>          | <i>n</i> | <i>M (SD)</i> | <i>n</i> |
| (t1)                         |                                |          |                        |          |               |          |
| Academic Self-Efficacy (ASE) | 4.11 (0.88)                    | 45       | 4.12 (0.92)            | 48       | 4.12 (0.90)   | 93       |
| Metacognitive LS (MCLS)      | 3.59 (0.53)                    | 45       | 3.79 (0.51)            | 48       | 3.70 (0.53)   | 93       |
| Experience with RW (ERW)     | 2.29 (1.39)                    | 45       | 2.06 (1.28)            | 48       | 2.17 (1.33)   | 93       |
| Attitude towards RW (ARW)    | 3.92 (0.66)                    | 45       | 3.94 (0.74)            | 48       | 3.93 (0.70)   | 93       |
| Reflection Score (RS1)       | 23.38 (4.69)                   | 45       | 19.31 (7.64)           | 48       | 21.28 (6.67)  | 93       |
| (t2)                         |                                |          |                        |          |               |          |
| Feedback Engagement (FE)     | 3.94 (0.56)                    | 43       | 3.91 (0.55)            | 41       | 3.92 (0.55)   | 84       |
| Reflection Score (RS2)       | 26.28 (6.59)                   | 43       | 21.55 (7.85)           | 42       | 23.94 (7.58)  | 85       |

Note. LS = learning strategies, RW = reflective writing.

**Table 3.** Correlation coefficients for reflection scores and reflection-related dispositions.

|           | ASE (t1) | MCLS (t1) | ERW (t1) | ARW (t1) | FE (t2) | RS (t1)  |
|-----------|----------|-----------|----------|----------|---------|----------|
| MCLS (t1) | *** 0.64 |           |          |          |         |          |
| ERW (t1)  | 0.03     | 0.03      |          |          |         |          |
| ARW (t1)  | * 0.25   | ** 0.27   | 0.15     |          |         |          |
| FE (t2)   | *** 0.38 | *** 0.43  | 0.07     | * 0.24   |         |          |
| RS (t1)   | * 0.24   | 0.20      | 0.09     | 0.06     | * 0.26  |          |
| RS (t2)   | ** 0.28  | 0.21      | * 0.23   | * 0.25   | ** 0.29 | *** 0.56 |

Note. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

## 5.2. Linear Mixed Effects Model

A Linear Mixed Effects Model was used to address both research questions. The full model (including fixed and random effects) explained 60.6% of the total variance in reflection scores (Conditional  $R^2$ ).

Regarding the fixed effects, time emerged as a significant predictor, with  $\hat{\beta} = 0.41$ , 95% CI [0.14, 0.68],  $t(81.90) = 3.02$ ,  $p = 0.003$ , indicating a general increase in reflection scores from t1 to t2 by nearly half a standard deviation. Feedback condition also reached significance in the model,  $\hat{\beta} = -0.42$ , 95% CI [-0.80, -0.04],  $t(125.82) = -2.17$ ,  $p = 0.032$ , with participants in the AI feedback group scoring, on average, nearly half a standard deviation lower than those in the instructor feedback group. However, the interaction term (RQ 1) was non-significant, suggesting that both feedback groups showed comparable gains in reflection scores from pre- to post-measurement.

Concerning reflection-related dispositions (RQ 2), only academic self-efficacy emerged as a significant positive predictor of reflection scores,  $\hat{\beta} = 0.25$ , 95% CI [0.05, 0.45],  $t(81.10) = 2.52$ ,  $p = 0.014$ . The fixed effects accounted for 23.7% of the variance in reflection scores (Marginal  $R^2$ ).

The intraclass correlation coefficient (ICC) was 0.48, indicating that a substantial proportion of variance in reflection scores was attributable to differences between participants (random effects).

## 6. Discussion

### 6.1. Summary

In accordance with the first research question (RQ 1), the main objective of this study was to analyze the effects of AI feedback on the reflection level of pre-service teachers, specifically compared to instructor feedback. The results indicate that pre-service teachers

in both groups slightly increased their reflection scores from the first to the second measurement point. This is consistent with previous findings and indicates that, on the one hand, reflection skills can be improved (e.g., [Von Aufschnaiter et al., 2019](#); [Häcker, 2019](#)) and, on the other hand, the provision of feedback may have a positive impact on this learning process ([Darling-Hammond & Bransford, 2007](#); [Jahncke et al., 2018](#); [Korthagen & Kessels, 1999](#); [Schellenbach-Zell et al., 2023](#)). Nonetheless, pre-service teachers' reflections generally remained on a lower level, consistent with previous findings (e.g., [Zhang et al., 2023, 2024](#)), highlighting that the ability to reflect (in writing) is a complex skill (e.g., [Aeppli & Lötscher, 2016](#)) that requires time and multiple interventions to develop (e.g., [Häcker, 2019](#)). Furthermore, other factors appear to determine the development of reflection skills (e.g., [Von Aufschnaiter et al., 2019](#)), the handling of feedback (e.g., [Deeva et al., 2021](#)), and the use of AI tools, as well as the acceptance of the output results ([Zhang et al., 2023](#)). Pre-service teachers who received AI feedback increased their reflection skills similarly to the lecturer feedback group. Previous research has shown that modern technologies such as AI and machine learning can be used to assess the quality of reflective writing ([Barthakur et al., 2022](#); [Jung et al., 2022](#); [Zhang et al., 2024](#)). Our findings add novel evidence that AI-based feedback, when guided by didactically informed prompts, can be at least as effective as feedback from experienced instructors. This suggests that, with future developments in this field (e.g., [Chong et al., 2020](#); [Fleckenstein et al., 2023](#); [Suvendu & Deb, 2024](#); [Wongvorachan et al., 2022](#); [Wulff et al., 2023](#)), the use of AI can positively impact learning processes, and its effectiveness may further increase. However, we also found that the group receiving feedback from instructors performed slightly better than the group receiving feedback from AI—even if this difference was not significant in our model. Therefore, it is certainly worth investigating whether, despite similar effects, a closer examination could reveal differences between the two forms of feedback. For example, despite uniform criteria and formal alignment, minor (qualitative) differences could exist, which may ultimately lead to minor differences in their impact on the development of (written) reflection skills. This could further indicate that the use of AI has limitations and, at least currently, does not (yet) achieve better results.

Regarding the second research question (RQ 2), we found evidence of a significant relationship between academic self-concept and the development of reflection skills, which aligns with existing research ([Stender et al., 2021](#)). However, other factors in the model did not prove to be significant predictors of the development of reflection skills. This is only partially consistent with existing findings: several empirical studies have reported relationships between prior experience with (written) reflections (e.g., [Fraij & Kirschner, 2017](#)) and attitude towards reflections (e.g., [Göbel & Neuber, 2020](#); [Stender et al., 2021](#)) and the development of reflective skills. Instead, our results suggest that individual differences between pre-service teachers exert a stronger influence on reflection levels, which is again in line with prior research (e.g., [Aeppli & Lötscher, 2016](#); [Lambrecht & Bosse, 2020](#); [Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg, LI Hamburg, 2020](#)) and indicates that individual learning prerequisites and learning arrangements play a significant role, especially in the development of reflective skills (e.g., [Saric & Steh, 2017](#)). Furthermore, the various theoretical models for mapping reflection processes (e.g., [Aeppli & Lötscher, 2016](#); [Von Aufschnaiter et al., 2019](#); [Blömeke et al., 2015](#)) already indicate that a multitude of (additional) factors influence the development of reflective skills (e.g., [Sudirman et al., 2024](#)).

## 6.2. Limitations

This study faces several limitations that warrant acknowledgment and consideration. Due to the complexity of the intervention, only a relatively small sample of fully

evaluated cases was available, and only a few covariates could be used for the analysis, with corresponding implications for the power and generalizability of the findings. Consequently, several known direct or indirect factors influencing the development of reflective skills (e.g., Aeppli & Lötscher, 2016) could not be fully considered in the analyses, such as subject-specific competencies (Henderson, 1992) and motivational or emotional learning prerequisites (Leonhard & Rihm, 2011). Furthermore, other constructs beyond metacognitive learning strategies have not been included (Möllers, 2014; Pintrich & Garcia, 1993). Regarding the opportunities and limitations of feedback concepts based on machine learning or AI, it should be noted that the aforementioned personal characteristics of learners also play a role in this context: factors such as AI acceptance or attitudes towards automated learning processes could influence the effectiveness of the feedback provided (e.g., Zhang et al., 2024). Additionally, no specific pedagogical concepts or implementation strategies targeting AI use were applied (e.g., Eden et al., 2024; Suvendu & Deb, 2024). Such concepts could potentially mitigate negative attitudes towards the use of AI in learning processes. In this context, it should also be noted that this study was conducted at a German university, and the results cannot be easily generalized to other countries with different teacher training structures. This limitation becomes even more pronounced if one were to extrapolate the results to other fields. In our design, the intervention only provided for a one-time provision of feedback. Feedback research has shown that multiple iterations are often necessary to realize the full potential of feedback (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Furthermore, to keep feedback conditions constant, specific attributes of the two feedback forms were not integrated into the intervention: for example, the near-immediate availability of AI feedback to learners (e.g., Deeva et al., 2021; Fleckenstein et al., 2023; Ullmann, 2015) or the ability of instructor feedback to consider previously unspecified criteria and respond more empathetically (e.g., O'Donovan et al., 2019; Ossenbeger et al., 2019) was deliberately not utilized. Studies using LLMs such as ChatGPT often face the problem of non-replicable results: despite using the same prompt at different times, different outputs may occur (e.g., Cao et al., 2025; Mondal et al., 2024). This ensures, on the one hand, that different results could potentially be obtained at a different study point in time; on the other hand, the study cannot be easily repeated or expanded. Another limitation concerns the assessment of reflections: reflective ability is often evaluated using stage models (e.g., Hatton & Smith, 1995). The small number of stages and the major changes often required to reach a higher level can obscure small developments (Zhang et al., 2023). For this reason, among others, the present study used a multi-layered criteria grid as the basis for assessment and feedback (e.g., Ullmann, 2019). However, although we applied an assessment method based on a theoretical and multi-layered criteria grid (e.g., Von Aufschnaiter et al., 2019), allowing for detailed ratings of different aspects of reflective ability, the criteria used may still limit the thoroughness of our evaluation. Therefore, the criteria grid and associated coding guidelines require further testing. Overall, it remains unclear whether AI and instructor feedback can be expected to have different effects on the various aspects of reflection skills, as only the overall score was included in the tested model. Additionally, little attention was paid to the form of feedback given to students, which research indicates can influence its effects (e.g., Hattie & Timperley, 2007). For more significant improvement in reflection skills (Aeppli & Lötscher, 2016; Argyris & Schön, 2024; Leonhard & Rihm, 2011), multiple measures addressing different aspects of reflection would be necessary (Artmann & Herzmann, 2016; Borko et al., 1997).

### 6.3. Implications and Future Directions

The study results show that pre-service teachers can develop their written reflection skills both through instructor feedback and, similarly, when receiving AI-generated feed-

back. This important finding may contribute to the further development of instructional quality in teacher education, offering pre-service teachers the opportunity to receive anonymous, criteria-based, and frequent feedback almost immediately, at nearly any time of day. Since reflections not only support learning processes in university courses in teacher training but also play a significant role in practical phases, their potential range of applications could be significantly expanded (e.g., Häcker, 2019). However, our results indicate that feedback alone is not sufficient to significantly improve the reflection skills of pre-service teachers. To achieve real improvement, several concepts must be used in combination: in addition to feedback, detailed explanations, prompts, examples, and the consideration of individual learning requirements and motivation are necessary (Artmann & Herzmann, 2016; Borko et al., 1997). Developing corresponding, more complex concepts requires further empirical insights. In this context, it is particularly important to examine the quality of feedback, both in general and in comparison, between the two actors. Specifically, qualitative analyses of the feedback provided could generate valuable insights in this area. Closely linked to these considerations is the idea of generating higher-quality feedback by varying the prompt or by improving prompt engineering and/or using powerful LLMs, thereby potentially enhancing the development of reflective skills. Furthermore, it should be noted that the integration of AI in learning environments offers research potential beyond technical developments; the ethical and moral implications of AI use must also increasingly be addressed in future research.

**Author Contributions:** Conceptualization, F.H., T.-M.D. and M.G.-Z.; Methodology, F.H. and T.-M.D.; Validation, F.H. and M.G.-Z.; Formal analysis, T.-M.D.; Investigation, F.H., T.-M.D. and M.G.-Z.; Data curation, T.-M.D. and L.P.; Writing—original draft preparation, F.H. and T.-M.D.; Writing—review and editing, F.H., T.-M.D., L.P. and M.G.-Z.; Visualization, F.H. and T.-M.D.; Supervision, F.H. and M.G.-Z.; Project administration, L.P. and M.G.-Z.; Funding acquisition, M.G.-Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Federal Ministry of Education and Research under grant number 16DHB4019 obtained by Prof. Dr. Michaela Gläser-Zikuda.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the Friedrich-Alexander-Universität Erlangen-Nürnberg (protocol code 20200831 01; 1 September 2020).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Due to the presence of personally identifiable information within the dataset, it is not publicly shareable in accordance with privacy protection laws and ethical guidelines of the universities Erlangen-Nürnberg and Berlin involved in this research.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## Appendix A. ChatGPT 4.0 Prompt for Reflective Writing Feedback

### 1. Original German prompt as used for providing feedback through ChatGPT 4.0

“Sie sind jetzt Dozentin oder Dozent im Bereich der Lehrerbildung. Sie unterrichten das Fach Schulpädagogik. Bitte geben Sie Ihren Studierenden eine Rückmeldung zu ihrem Reflexionsschreiben. Dabei müssen Sie die folgenden Regeln beachten:

1. Bitte bewerten Sie anhand der untenstehenden Kriterien.
2. Bitte stellen Sie Feedback in Textform dar (die Kriterien sollen nicht im Text erscheinen).
3. Bitte achten Sie auf Höflichkeit im Ausdruck, Kontinuität und Lesbarkeit der Worte.

- A Formale Gestaltung
  - (1) Struktur
  - (2) Umfang
  - (3) Sprache
- B Inhalt
  - (4) Reflexionsausrichtung:
    - (4.1) Angabe eines Reflexionsziels
    - (4.2) Begründung des Reflexionsziels
  - (5) Bezug zur Unterrichtssituation
  - (6) Multiperspektivität:
    - (6.1) Objektiv-fachliche Perspektive
    - (6.2) Subjektiv-persönliche Perspektive
    - (6.3) Zeitliche Perspektive
    - (6.4) Weitere Perspektiven (z.B. aus dem Blickwinkel unterschiedlicher Personen, zu unterschiedlichen Zeitpunkten usw.)
  - (7) Gedankliche Verknüpfung der unterschiedlichen Perspektiven und Argumente
  - (8) Weiterführende Gedanken
- C Gesamteinschätzung hinsichtlich der Reflexionstiefe"

## 2. Translated prompt (English)

"You are now a university instructor in the field of teacher education. You are teaching the subject of school pedagogy. Please provide your students with feedback on their reflective writing. In doing so, you must adhere to the following rules:

1. Please evaluate based on the criteria listed below.
  2. Please provide your feedback in continuous text form (the criteria themselves should not appear explicitly in the text).
  3. Please ensure polite expression, coherence, and readability of your wording.
- A Formal Aspects
    - (1) Structure
    - (2) Length
    - (3) Language
  - B Content
    - (4) Reflective Orientation:
      - (4.1) Specification of a reflection goal
      - (4.2) Justification of the reflection goal
    - (5) Reference to the teaching situation
    - (6) Multi-perspectivity:
      - (6.1) Objective-professional perspective
      - (6.2) Subjective-personal perspective
      - (6.3) Temporal perspective
      - (6.4) Additional perspectives (e.g., from the viewpoint of different individuals, at different points in time, etc.)
    - (7) Conceptual integration of the different perspectives and arguments
    - (8) Forward-looking thoughts

## C Overall assessment of the depth of reflection"

## References

- Abdalina, L., Bulatova, E., Gosteva, S., Kunakovskaya, L., & Frolova, O. (2022). Professional development of teachers in the context of the lifelong learning model: The role of modern technologies. *World Journal on Educational Technology*, 14(1), 117–134. [\[CrossRef\]](#)
- Aeppli, J., & Lötscher, H. (2016). EDAMA—Ein Rahmenmodell für Reflexion. *Beiträge zur Lehrerinnen und Lehrerbildung*, 34(1), 78–97. [\[CrossRef\]](#)
- Albert, S. (2016). Die Bedeutung der reflexiven Selbstforschung für die Professionalisierung von Lehrpersonen. *Haushalt in Bildung & Forschung*, 5(4), 35–46.
- Argyris, C., & Schön, D. A. (2024). *Die lernende Organisation. Grundlagen, Methode, Praxis* (3rd ed.). Klett-Cotta.
- Artmann, M., & Herzmann, P. (2016). Portfolioarbeit im Urteil der Studierenden—Ergebnisse einer Interviewstudie zur Lehrerinnenbildung im Kölner Modellkolleg. In S. Ziegelbauer, & M. Gläser-Zikuda (Eds.), *Portfolio als Innovation in Schule, Hochschule und Lehrerinnenbildung* (pp. 131–146). Klinkhardt.
- Barthakur, A., Joksimovic, S., Kovanovic, V., Mello, R. F., Taylor, M., Richey, M., & Pardo, A. (2022). Understanding depth of reflective writing in workplace learning assessments using machine learning classification. *IEEE Transactions on Learning Technologies*, 15(5), 567–578. [\[CrossRef\]](#)
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. [\[CrossRef\]](#)
- Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, 54(5), 1160–1173. [\[CrossRef\]](#)
- Benade, L. (2017). *Being a teacher in the 21st century: A critical New Zealand research study*. Springer. [\[CrossRef\]](#)
- Berger, J. (2022). *Selbstreflexion im Studium: Entwicklung und Validierung eines Messinstruments bei Lehramtsstudierenden*. Technische Universität Darmstadt. [\[CrossRef\]](#)
- Blömeke, S., Gustafson, J.-E., & Shavelson, R. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie*, 123(1), 3–13. [\[CrossRef\]](#)
- Borko, H., Michalec, P., Timmons, M., & Siddle, J. (1997). Student teaching portfolios: A tool for promoting reflective practice. *Journal of Teacher Education*, 48(5), 345–357. [\[CrossRef\]](#)
- Boyd, E. M., & Fales, A. W. (1983). Reflective learning: Key to learning from experience. *Journal of Humanistic Psychology*, 23(2), 99–117. [\[CrossRef\]](#)
- Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, 23(3), 389–411. [\[CrossRef\]](#)
- Buckingham Shum, S., Lim, L.-A., Boud, D., Bearman, M., & Dawson, P. (2023). A comparative analysis of the skilled use of automated feedback tools through the lens of teacher feedback literacy. *International Journal of Educational Technology in Higher Education*, 20(1), 40. [\[CrossRef\]](#)
- Cao, J., Li, M., Wen, M., & Cheung, S.-C. (2025). A study on prompt design, advantages and limitations of ChatGPT for deep learning program repair. *Automated Software Engineering*, 32(1), 30. [\[CrossRef\]](#)
- Chan, C. K., & Lee, K. K. (2021). Reflection literacy: A multilevel perspective on the challenges of using reflections in higher education through a comprehensive literature review. *Educational Research Review*, 32, 100376. [\[CrossRef\]](#)
- Chang, D. H., Lin, M. P.-C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using ai chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, 15(17), 12921. [\[CrossRef\]](#)
- Chong, C., Sheikh, U. U., Samah, N. A., & Sha'ameri, A. Z. (2020). Analysis on reflective writing using natural language processing and sentiment analysis. *IOP Conference Series: Materials Science and Engineering*, 884(1), 12069. [\[CrossRef\]](#)
- Darling-Hammond, L., & Bransford, J. (2007). *Preparing teachers for a changing world: What teachers should learn and be able to do*. John Wiley & Sons.
- Deeva, G., Bogdanova, D., Serral, E., Snoeck, M., & de Weerd, J. (2021). A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, 162, 104094. [\[CrossRef\]](#)
- Desautel, D. (2009). Becoming a thinking thinker: Metacognition, self-reflection, and classroom practice. *Teachers College Record*, 111(8), 1997–2020. [\[CrossRef\]](#)
- Dewey, J. (2022). *How we think: A restatement of the relation of reflective thinking to the educative process* (Reprint). DigiCat. (Original work published 1910).
- Eden, C. A., Chisom, O. N., & Adeniyi, I. S. (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. *Magna Scientia. Advanced Research and Reviews*, 10(2), 6–13. [\[CrossRef\]](#)

- Feder, L., Fütterer, T., & Cramer, C. (2021). Einstellungen Studierender zur Portfolioarbeit. Theoriebasierte Erfassung und erste deskriptive Befunde. In N. Beck, T. Bohl, & S. Meissner (Eds.), *Vielfältig herausgefordert. Forschungs- und Entwicklungsfelder der Lehrerbildung auf dem Prüfstand. Diskurse und Ergebnisse der ersten Förderphase der Qualitätsoffensive Lehrerbildung an der Tübingen school of education* (pp. 209–221). Tübingen University Press. [CrossRef]
- Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6, 1162454. [CrossRef] [PubMed]
- Fraij, A., & Kirschner, S. (2017, September 25–27.). *Die Messung reflexionsbezogenen Wissens von Studierenden des Lehramts und der Erziehungswissenschaft [conference paper]*. Tagung der Sektion "Empirische Bildungsforschung—educational research and governance" der Arbeitsgruppe für empirische Pädagogische Forschung (AEPF), Tübingen, Germany.
- George, A. S. (2023). Preparing students for an ai-driven world: Rethinking curriculum and pedagogy in the age of artificial intelligence. *Partners Universal Innovative Research Publications (PUIRP)*, 1(2), 112–136. [CrossRef]
- Gibbs, G. (1988). *Learning by doing: A guide to teaching and learning methods*. Oxford University Press.
- Gibson, A., Kitto, K., & Bruza, P. (2016). Towards the discovery of learner metacognition from reflective writing. *Journal of Learning Analytics*, 3(2), 22–36. [CrossRef]
- Gore, J. M., & Zeichner, K. M. (1991). Action research and reflective teaching in preservice teacher education: A case study from the United States. *Teaching and Teacher Education*, 7(2), 119–136. [CrossRef]
- Göbel, K., & Neuber, K. (2020). Einstellungen zur Reflexion von angehenden und praktizierenden Lehrkräften. *Empirische Pädagogik*, 34(1), 64–78.
- Hansen, C., & Rachbauer, T. (2018). *Reflektieren? Worauf und wozu? Arbeiten mit dem E-Portfolio—Ein Reflexionsinstrument für die LehrerInnenbildung am Beispiel der Universität Passau*. E-teaching.org. Available online: [https://www.e-teaching.org/etresources/pdf/erfahrungsbericht\\_2018\\_hansen\\_rachbauer\\_arbeiten\\_mit\\_dem\\_e\\_portfolio\\_reflexionsinstrument\\_fuer\\_die\\_lehrerbildung.pdf](https://www.e-teaching.org/etresources/pdf/erfahrungsbericht_2018_hansen_rachbauer_arbeiten_mit_dem_e_portfolio_reflexionsinstrument_fuer_die_lehrerbildung.pdf) (accessed on 26 June 2025).
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. [CrossRef]
- Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, 11(1), 33–49. [CrossRef]
- Häcker, T. (2019). Reflexive Professionalisierung. Anmerkungen zu dem ambitionierten Anspruch, die Reflexionskompetenz angehenden Lehrkräfte umfassend zu fördern. In M. Degeling, N. Franken, S. Freund, S. Greiten, D. Neuhaus, & J. Schellenbach-Zell (Eds.), *Herausforderung Kohärenz: Praxisphasen in der universitären Lehrerbildung. Bildungswissenschaftliche und fachdidaktische Perspektiven* (pp. 81–96). Klinkhardt.
- Henderson, J. G. (1992). *Reflective teaching: Becoming an inquiring educator*. MacMillan.
- Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers*, 16(2), 2577–2587. [CrossRef]
- Idel, T.-S., Schütz, A., & Thünemann, S. (2020). Professionalität im Handlungsfeld Schule. In J. Dinkelaker, K.-U. Hugger, T.-S. Idel, A. Schütz, & S. Thünemann (Eds.), *Professionalität und Professionalisierung in pädagogischen Handlungsfeldern: Schule, Medienpädagogik, Erwachsenenbildung* (pp. 13–82). Barbara Budrich.
- Jaeger, E. L. (2013). Teacher reflection: Supports, barriers, and results. *Issues in Teacher Education*, 22(1), 89–104.
- Jahncke, H., Berding, F., Porath, J., & Magh, K. (2018). Einfluss von Feedback auf die (Selbst-)Reflexion von Lehramtsstudierenden. *Die Hochschullehre*, 4(1), 505–530. [CrossRef]
- Jones, J. L., & Jones, K. A. (2013). Teaching reflective practice: Implementation in the teacher-education setting. *The Teacher Educator*, 48(1), 73–85. [CrossRef]
- Jung, Y., Wise, A. F., & Allen, K. L. (2022). Using theory-informed data science methods to trace the quality of dental student reflections over time. *Advances in Health Sciences Education: Theory and Practice*, 27(1), 23–48. [CrossRef]
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. [CrossRef]
- KMK—Ständige Konferenz der Kultusminister der Bundesrepublik Deutschland. (2022). *Standards für die Lehrerbildung: Bildungswissenschaften* (= Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 16.05.2019). Sekretariat der Kultusministerkonferenz. Available online: [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf) (accessed on 23 June 2025).
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163. [CrossRef]
- Korthagen, F. A. J. (1999). Linking reflection and technical competence: The logbook as an instrument in teacher education. *European Journal of Teacher Education*, 22(2/3), 191–207. [CrossRef]
- Korthagen, F. A. J. (2018). Making teacher education relevant for practice: The pedagogy of realistic teacher education. *ORBIS SCHOLAE*, 5(2), 31–50. [CrossRef]

- Korthagen, F. A. J., & Kessels, J. P. (1999). Linking theory and practice: Changing the pedagogy of teacher education. *Educational Researcher*, 28(4), 4–17. [CrossRef]
- Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., & Dawson, S. (2018, March 5–9). *Understand students' self-reflections through learning analytics*. 8th International Conference on Learning Analytics and Knowledge (pp. 389–398), Sydney, Australia. [CrossRef]
- Körkkö, M., Kyrö-Ämmälä, O., & Turunen, T. (2016). Professional development through reflection in teacher education. *Teaching and Teacher Education*, 55, 198–206. [CrossRef]
- Lambrecht, J., & Bosse, S. (2020). Lässt sich die Reflexionsfähigkeit von angehenden Lehrkräften verändern? *HLZ, Herausforderung Lehrer\*Innenbildung Zeitschrift zur Konzeption, Gestaltung und Diskussion*, 3(2), 137–150. [CrossRef]
- Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg, LI Hamburg. (2020). *Reflexionskompetenz fördern. Reflexion und Reflexionskompetenz in der Lehrkräftebildung*. A & C Druck und & Verlag Hamburg.
- Lenske, G., & Lohse-Bossenz, H. (2023). Stichwort: Reflexion im pädagogischen Kontext. *Zeitschrift für Erziehungswissenschaft*, 26(5), 1133–1164. [CrossRef]
- Leonhard, T., & Rihm, T. (2011). Erhöhung der Reflexionskompetenz durch Begleitveranstaltungen zum Schulpraktikum? Konzeption und Ergebnisse eines Pilotprojekts mit Lehramtsstudierenden. *Lehrerbildung auf dem Prüfstand*, 4(2), 240–270.
- Li, L., & Kim, M. (2024). It is like a friend to me: Critical usage of automated feedback systems by self-regulating English learners in higher education. *Australasian Journal of Educational Technology*, 40(1), 1–18. [CrossRef]
- Meyer-Siever, K., & Levin, A. (2018). Entwicklung der Reflexionskompetenz im Rahmen eines fächerübergreifenden E-portfolios. *Resonanz, Magazin für Lehre und Studium an der Universität Bremen*, 2018(1), 24–31.
- Mondal, S., Bappon, S. D., & Roy, C. K. (2024, April 15–16). *Enhancing user interaction in ChatGPT: Characterizing and consolidating multiple prompts for issue resolution*. 21st International Mining Software Repositories (MSR '24) (pp. 222–226), Lisbon, Portugal. [CrossRef]
- Möllers, L. (2014). Reflexionskompetenz und Innovationskompetenz im Berufsfeldpraktikum. In A. Schöning, M. Heer, M. Pahl, F. Diehr, E. Parusel, A. Tinnefeld, & J. Walke (Eds.), *Das Berufsfeld-praktikum als Professionalisierungselement*. Grundlagen, Konzepte, Beispiele für das Lehramtsstudium (pp. 166–172). Klinkhardt. [CrossRef]
- Nehyba, J., & Štefánik, M. (2023). Applications of deep language models for reflective writings. *Education and Information Technologies*, 28(3), 2961–2999. [CrossRef]
- O'Donovan, B. M., den Outer, B., Price, M., & Lloyd, A. (2019). What makes good feedback good? *Studies in Higher Education*, 46(2), 318–329. [CrossRef]
- Ossenberg, C., Henderson, A., & Mitchell, M. (2019). What attributes guide best practice for effective feedback? A scoping review. *Advances in Health Science Education*, 24(2), 383–401. [CrossRef]
- Philipp, J. (2023). Reflexionsfähigkeit in der interdisziplinären Lehre. Hochschuldidaktische Perspektiven auf Lernziele und Prüfungen. In M. Braßler, S. Brandstädter, & S. Lerch (Eds.), *Interdisziplinarität in der Hochschullehre* (=Interdisziplinäre Lehre, Bd. 1) (pp. 149–161). wbv Publikation. [CrossRef]
- Pintrich, P. R., & Garcia, T. (1993). Intraindividual differences in students' motivation and self-regulated learning. *Zeitschrift für Pädagogische Psychologie*, 7(3), 99–107.
- Poldner, E., van der Schaaf, M., Simons, P. R.-J., van Tartwijk, J., & Wijngaards, G. (2014). Assessing student teachers reflective writing through quantitative content analysis. *European Journal of Teacher Education*, 37(3), 348–373. [CrossRef]
- R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Revelle, R. (2025). *psych: Procedures for psychological, psychometric, and personality research* (R package version 2.5.6). Northwestern University. Available online: <https://CRAN.R-project.org/package=psych> (accessed on 26 June 2025).
- Riel, M. (2022). *Empirische Überprüfung eines verfahrens zur messung von reflexion bei lehramtsstudierenden*. Universität Passau.
- Ross, M., Bohlmann, J., & Marren, A. (2024). Reflective writing as summative assessment in higher education: A systematic review. *Journal of Perspectives in Applied Academic Practice*, 12(1), 54–67. [CrossRef]
- Russell, T., & Korthagen, F. (Eds.). (2013). *Teachers who teach teachers: Reflections on teacher education*. Routledge.
- Ryan, M. (2013). The pedagogical balancing act: Teaching reflection in higher education. *Teaching in Higher Education*, 18(2), 144–155. [CrossRef]
- Saric, M., & Steh, B. (2017). Critical reflection in the professional development of teachers: Challenges and possibilities. *CEPS Journal*, 7(3), 67–85. [CrossRef]
- Schellenbach-Zell, J., Molitor, A. L., Kindlinger, M., Trempler, K., & Hartmann, U. (2023). Wie gelingt die Anregung von Reflexion über pädagogische Situationen unter Nutzung bildungswissenschaftlicher Wissensbestände? Die Bedeutung von Prompts und Feedback. *Zeitschrift für Erziehungswissenschaft*, 26(5), 1189–1211. [CrossRef]
- Schiefele, U., Moschner, B., & Husstegge, R. (2002). *Skalenhandbuch SMILE-Projekt* (unveröffentlichtes Manuskript). Universität, Abteilung für Psychologie.
- Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

- Sebastian, G. (2023). Privacy and data protection in ChatGPT and other AI chatbots. *International Journal of Security and Privacy in Pervasive Computing*, 15(1), 1–14. [CrossRef]
- Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L. (2022). A review of the trends and challenges in adopting natural language processing methods for education feedback analysis. *Institute of Electrical and Electronics Engineers (IEEE) Access*, 10(2), 56720–56739. [CrossRef]
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. [CrossRef]
- Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open*, 6(1), 100169. [CrossRef]
- Stender, J., Watson, C., Vogelsang, C., & Schaper, N. (2021). Wie hängen bildungswissenschaftliches Professionswissen, Einstellungen zu Reflexion und die Reflexionsperformanz angehender Lehrpersonen zusammen? *HLZ, Herausforderung Lehrer\*Innenbildung*, 4(1), 229–248. [CrossRef]
- Sudirman, A., Gemilang, A. V., Kristanto, T. M. A., Robiasih, R. H., Hikmah, I., Nugroho, A. D., Karjono, J. C. S., Lestari, T., Widayarni, T. L., Prastanti, A. D., Susanto, M. R., & Rais, B. (2024). Reinforcing reflective practice through reflective writing in higher education: A systematic review. *International Journal of Learning, Teaching and Educational Reserch*, 23(5), 454–474. [CrossRef]
- Suvendu, R., & Deb, P. S. (2024). AI-driven flipped classroom: Revolutionizing education through digital pedagogy. *British Journal of Education, Learning and Development Psychology*, 7(2), 169–179. [CrossRef]
- Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education*, 23(1), 1–23. Available online: <https://www.learntechlib.org/primary/p/222408/> (accessed on 21 June 2025).
- Ullmann, T. D. (2015). *Automated detection of reflection in texts: A machine learning based approach* [Ph.D. thesis, The Open University]. [CrossRef]
- Ullmann, T. D. (2019). Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education*, 29(1), 217–257. [CrossRef]
- Van den Berg, G., & Du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences*, 13(10), 998. [CrossRef]
- Von Aufschnaiter, C., Fraij, A., & Kost, D. (2019). Reflexion und Reflexivität in der Lehrerbildung. *HLZ, Herausforderung Lehrer\*Innenbildung*, 2(1), 144–159. [CrossRef]
- Watini, S., Davies, G., & Andersen, N. (2024). Cybersecurity in learning systems: Data protection and privacy in educational information systems and digital learning environments. *International Transactions on Education Technology (ITEE)*, 3(1), 26–35. [CrossRef]
- Wiese, L. J., Patil, I., Schiff, D. S., & Magana, A. J. (2025). AI ethics education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 8(3), 100405. [CrossRef]
- Wild, K.-P., & Schiefele, U. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 15(4), 185–200.
- Winkel, M. (2025). Society in charge: The connection of artificial intelligence, responsibility, and ethics in German media discourse. *AI Ethics*, 5(3), 2839–2866. [CrossRef]
- Wongvorachan, T., Lai, K. W., Bulut, O., Tsai, Y.-S., & Chen, G. (2022). Artificial intelligence: Transforming the future of feedback in education. *Journal of Applied Testing Technology*, 23(1), 95–116. Available online: <https://jattjournal.net/index.php/atp/article/view/170387> (accessed on 19 June 2025).
- Wulff, P., Mientus, L., Nowak, A., & Borowski, A. (2023). Utilizing a pretrained language model (BERT) to classify preservice physics teachers' written reflections. *International Journal of Artificial Intelligence in Education*, 33(3), 439–466. [CrossRef]
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 2021, 8812542. [CrossRef]
- Zhang, C., Hofmann, F., Plößl, L., & Gläser-Zikuda, M. (2024). Classification of reflective writing: A comparative analysis with shallow machine learning and pre-trained language models. *Education and Information Technologies*, 29(16), 21593–21619. [CrossRef]
- Zhang, C., Schießl, J., Plößl, L., Hofmann, F., & Gläser-Zikuda, M. (2023). Evaluating reflective writing in pre-service teachers: The potential of a mixed-methods approach. *Education Sciences*, 13(12), 12134. [CrossRef]
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.