Castledown

OPEN ACCESS

Technology in Language Teaching & Learning
ISSN 2652-1687
https://www.castledown.com/journals/tltl/

*Technology in Language Teaching & Learning, 8,* 103803 (2026)
https://doi.org/10.29140/tltl.2026.103803

# Transforming Teacher Training: The Power of ChatGPT in Lesson Practice and Reflection

ALI AL GHAITHI[a]
BEHNAM BEHFOROUZ[b]
ABDULLAH AL MAQBALI[c]

[a]*Sohar University, Oman*
AGhaithi@su.edu.om

[b]*University of Technology and Applied Sciences, Shinas, Oman*
Behnam.Behforouz@utas.edu.om

[c]*Sohar University, Oman*
AMmaqbali@su.edu.om

## Abstract

This research investigated the impact of ChatGPT on pre-service teachers' evaluations through classroom demonstrations and written reflections on lesson planning. The research involved 40 Omani pre-service teachers, who were randomly assigned to either the experimental or the control group, with 20 participants in each group. The experimental group received in-person instruction and essential training similar to the control group, but they also used ChatGPT as a facilitator during their preparatory work. The experimental group achieved better results than the control group across all subsequent assessments that evaluated teaching demonstration and reflective report skills. The research findings provide essential value to teachers, institutions, and curriculum developers.

*Keywords:* ChatGPT; pre-service teachers; teaching practice; reflection

## Introduction

The method of gaining knowledge via experience to acquire new perspectives for future actions is generally called reflective practice (Boud et al., 1985). Reflective writing is one of the most critical

tools for new teachers to use in their reflection process. There are numerous critical advantages to the written reflection practice of pre-service instructors. Krol (1996) emphasizes that reflective writing is a "method that encourages reflection and is a valuable tool for facilitating dialogue between students and teachers" (p. 1). Additionally, reflective writing can serve as a self-evaluation tool for assessing one's perspective (Hume, 2009). In general, this method assists new teachers in forming their identities as educators and offers valuable insights into the instructional and educational activities of the student teachers (Whipp et al., 1997). General types of reflective writing include reflective or instructional journals (Bain et al., 2002; Ukrop et al., 2019), portfolios (Zeichner & Wray, 2001), blogs (Stiler & Philleo, 2003), and learning connections (Cardenas, 2014). However, Alger (2006) shows reflective activities produce reflection, but there is no scientific proof that they lead to better teacher practices. In addition, Mena-Marcos et al. (2013) showed that professionals build their expertise through reflective practice, yet this approach is rarely used in practice. Student reflective judgment develops differently based on age; however, most higher education studies fail to incorporate developmental psychological principles (King & Kitchener, 2004).

Leonhard and Rihm (2011) demonstrate that reflection serves as a vital link between academic learning and practical application, combining theoretical knowledge with educational methods. New teachers and pre-service educators need to learn integration methods because they struggle to apply their knowledge in actual classroom environments (Nguyen et al., 2014). Pre-service educators can identify their teaching strengths and weaknesses through systematic reflection (Rogers, 2001). Additionally, pre-service teachers who practice systematic reflection develop a better understanding of the teaching difficulties that accompany their educational responsibilities (Suphasri & Sumalee, 2021; Yu & Chiu, 2019). This method enables teachers to develop self-awareness and deepen their reflective practice (Korucu Kis & Kartal, 2019; McCormack, 2001). However, the development of pre-service instructors' reflective abilities encounters various obstacles (Azimi et al., 2019; Körkkö et al., 2016). The reflective process generates descriptive information that does not progress into analytical assessments. Delivering timely, effective, and constructive feedback to learners becomes difficult due to time constraints and professional duties (Ullmann, 2015).

To address issues in teaching evaluations and reflections, educators use artificial intelligence tools, including ChatGPT, to enhance their professional development and reflective practice (Behforouz et al., 2026; Behforouz & Al Ghaithi, 2025; Awidi, 2024). The ChatGPT reflection tool operates differently from standard reflection tools because it produces an experienced educator's perspective, which helps teachers perform detailed evaluations of their teaching practices and educational methods (ElSayary, 2024).

Research on ChatGPT indicates potential for the development of teacher reflection, but researchers must address various unanswered questions about artificial intelligence in educational settings (Deng & Lin, 2022; Griche & Bennis, 2026), but it lacks studies about how AI-generated visual materials affect new teachers' reflective practices (Cai et al., 2024). Additionally, the exact methods by which artificial intelligence systems can improve teaching, collaboration, and reflection practices remain unclear, despite artificial intelligence showing promise for educational development (Floris, 2023; Yu, 2024). Arefian et al. (2024) highlight this knowledge gap, noting that researchers recognize ChatGPT and other generative AI technologies affect learning, yet they need to study their potential for pre-service teacher professional development. The lack of thorough research on AI-based reflective practice enhancement requires immediate investigation, as it creates an urgent need to examine AI integration with teacher professional development. Therefore, the present investigation seeks to examine the efficacy of AI assistance in enhancing the quality of reflection and the instructional techniques of Omani pre-service teachers who use ChatGPT in their reflective work, by contrasting conventional reflection methods with ChatGPT-assisted reflection. The present study fills a substantial

void in literature, as no prior studies have investigated the relationship between ChatGPT and written reflection in pre-service teacher training. The results will improve comprehension of how to effectively integrate AI resources into teacher training courses to promote reflective practice. Therefore, the study attempts to find out the answers to the following research questions:

1. How does using a chatbot enhance the teaching practice of Omani pre-service teachers?
2. How does ChatGPT influence the quality and structure of written reflection reports among Omani pre-service teachers?

## Literature Review

The research by Karataş and Yüce (2024) analyzed 141 pre-service educators' opinions about artificial intelligence through their assessment of ChatGPT. The three-week Zoom-based study investigated how participants used AI for their cognitive work and humanistic and behavioral learning needs. The study showed that using ChatGPT in education transformed pre-service teachers' teaching methods, making learning more effective and accessible, according to Strauss and Corbin's coding system. Teachers can create individualized learning plans for students through this method, which simplifies their work while delivering tailored educational materials. The study limitations include reliance on teacher reflections as the data for analysis, a limited three-week timeframe that makes it challenging to draw strong conclusions about pedagogical change, and a focus on a single institution, which may limit the generalizability of the findings to other educational contexts.

The research by Luzano (2024) investigated how AI chatbots, including Gemini, Perplexity, and ChatGPT, affect mathematics education among 90 pre-service teachers at Bukidnon State University in the Philippines. The research results showed that pre-service teachers use ChatGPT to change their teaching methods, making mathematics education more accessible, interactive, and effective. The system generates higher-quality feedback, enabling customized learning experiences that help students build their self-confidence. The system enabled teachers to develop adaptable teaching approaches, leading to higher student satisfaction with the content. The educational practice of teaching has evolved through the implementation of ChatGPT, which provides vital support to boost student achievement in contemporary digital learning environments. Strengths were instrument reliability, a thorough multidimensional assessment of the activity, and solid statistics. The study design, however, was descriptive, preventing any conclusions regarding causality. Subjective perceptions of improved learning do not always translate to objectively measured learning gains. The study is also limited to a single institutional context and may not be generalized to other educational systems.

In another study, Quan et al. (2024) used a mixed-methods approach to examine ChatGPT use among 57 pre-service teachers who learned within a technologically enriched learning (TEL) framework. The research findings showed that pre-service teachers created better TEL artifacts by combining ChatGPT with prompts, thereby improving their teaching approaches. The design phase became more efficient because ChatGPT reduced extrinsic cognitive load, indicating better educational processes. The study's mixed-methods approach is a strength, as it combines quantitative and qualitative data to provide a more holistic understanding of the role of ChatGPT and prompt quality in technology-enhanced learning design. Practical implications include developing well-structured prompts when using ChatGPT to improve instruction, thereby supporting future instructional design efforts in this area. However, the context-specific sample and short-term follow-ups limit the generalizability of the findings and do not allow conclusions about the long-term impact on professional development.

Hashem et al. (2024) studied how ChatGPT generates customized educational materials for mathematics, science, and English language classes. The instructional framework was used to assess how

ChatGPT generated lesson plans through its responses. The results showed that ChatGPT demonstrates strong potential to assist teachers with lesson planning, content development, and instructional tasks when users provide high-quality prompts. The system produces better results through continuous user feedback, which allows teachers to create customized educational materials while working less. The study's greatest strength is its design, which enabled comparisons of lesson plans generated by generative AI across different subject areas. The study was more focused on the lesson plans generated by the generative AI than on their implementation and effects on students' learning in the classroom. Furthermore, efficacy was a function of prompt quality, suggesting that teacher digital literacy played a key mediating role in the effective integration of ChatGPT into classrooms.

## Method

### Participants

The study included 40 Omani pre-service teachers enrolled in the Teacher Preparation Diploma Course. The participants' ages ranged from 22 to 25 years, and they spoke Arabic as their native language. The program exists to train people with undergraduate degrees in literature or language to become English teachers at educational institutions. The program teaches candidates essential educational principles and teaching methods that form the basis of educational practice. The one-year program consists of two semesters, totaling its duration. The researchers used random selection to distribute 20 participants into each group from the pre-service instructor pool. The two courses included both males and females.

### Instruments

#### *Reflection Scale*

The researchers used the university reflection assessment, which forms part of the institutional curriculum. The written reflection grading consists of five main assessment areas, which together account for 20 points out of 100 marks. The lesson description section evaluates how well learning objectives, instructional methods, and classroom dynamics are presented. The critical analysis section assesses students' ability to demonstrate their comprehension of learners' engagement and teaching effectiveness. The reflection and personal comprehension section evaluates the ability to connect classroom observations with teaching principles and professional development. The evaluation of language and structure examines how well the writing is precise and coherent while reaching the 600–700-word length. The references and APA style section of the guidelines follows APA 7th edition formatting to evaluate proper citation and formatting. The evaluation framework for reflection papers includes a scale that assesses each criterion from outstanding to substandard performance. Multiple critical elements need thorough examination to develop an outstanding instruction reflection. The reflection begins with an extensive evaluation of the first microteaching session, which examines all activities, learning objectives, teaching methods, and classroom communication between teachers and students. The descriptive section should lead to a critical evaluation that uses specific examples from the observation to determine how well teaching methods align with learning objectives. The reflection demonstrates understanding of professional development by assessing needed skills and explaining how observed behaviors connect to teaching principles.

#### Demo Class Scale for the First and Second Assessment

The evaluation of instructional demonstration and lesson plan assessment will award a maximum of 120 points through 25 specific criteria, which fall under nine primary assessment areas, with each criterion receiving up to 5 marks. The assessment criteria include a well-written lesson plan (50 points),

the implementation of the lesson plan (50 points), and language, mechanics, and academic integrity (25 points). The weighting of the sections in the final mark is 40% for the written plan, 40% for the implementation of the plan, and 20% for language and integrity while presenting the plan. The evaluation system uses a 5-point scale that ranges from Exceptional (5) to Fail (0) to assess candidates, with Fail indicating failure of the basic requirements and Exceptional indicating superior mastery and creative understanding. The assessment system evaluates all aspects of teaching through written lesson plan assessment and direct observation of teaching methods.

### Assessment 1 and Assessment 2

The first assessment requires students to create lesson plans for students in grades 1–9. Students need to create a complete lesson plan using the template, adding learning objectives, potential challenges, instructional methods, materials, and assessment techniques. The assessment format follows a basic teaching approach, which asks student teachers to handle fundamental educational tasks while keeping their students involved. The assessment period lasts for 15 minutes. The short duration of the assessment requires students to demonstrate their mastery of fundamental teaching methods by effectively using suitable language for young students.

The assessment level increases in Assessment 2 because it focuses on teaching methods for students in upper high school grades (Grades 10–12). The educational approach needs to change for teaching older students because it must handle their complex mental and language abilities while using the complete lesson plan structure. Students can study subjects in detail and complete their challenging language assignments because the practice time extends to 20–25 minutes. The assessment provides teachers with methods for teaching university students to develop advanced language skills and creative thinking.

### Procedure

The research was conducted at one of the higher education universities in the Sultanate of Oman during the first semester of 2025, after obtaining ethical approval and participant consent. The research period spanned fourteen weeks.

The investigators taught both groups in person during the first 3 weeks, focusing on creating comprehensive lesson plans for various levels of English language proficiency. Six 120-minute instruction sessions were recorded and followed the same delivery patterns to ensure group uniformity. Learners in both groups used the university approach to implement in the micro-teaching subject, and both groups received direct training on how to teach the lesson plans using various teaching techniques. The sessions maintained a 70:30 ratio of theoretical to practical material, emphasizing practical work over theory. To establish formal training procedures, the investigator held a training session at the start of week 4 for both groups on how to write reflections for the demo class, utilizing the teacher's needs and instructions.

To prepare a lesson plan, the investigator asked both classes to choose any subject from any grade book published by the Ministry of Education of Oman. All learners from both groups presented their lessons during the microteaching hours after week four. The control and experimental groups were assessed using the standardized micro-teaching evaluation measure in separate classrooms by two different instructors to preserve assessment reliability. In accordance with the reflection regulations, the instructor asked both class participants to compose their first reflection at the start of week five. The instructor gave learners three hours to submit their reflections. Turnitin verified every submission and held one-on-one discussions to ensure the work was real. After week five, the instructor asked each pair to use the student-written comments to share their thoughts and gain knowledge about their demo courses. The instructors then provided feedback to learners in both groups based on the lesson plans and the demonstration class.

The first evaluation for both groups was carried out in week six of the systematic evaluation protocol's implementation. Every student was assigned a subject during class and given four hours to create a lesson plan. Learners were allowed to make little adjustments to the lesson plans after the instructor signed them, but not significant ones to preserve the integrity of the assessment. Every change was monitored and recorded. For comparison, students turned in paper copies to the instructor and digital versions via Turnitin, along with the initial drafts from class. The instructor asked students to create their second reflection using the same process and submit it during week 10. The researchers maintained identical submission requirements to create equal assessment conditions for both reflection assignments.

The experimental group received additional guidance about using ChatGPT to help them create their reflections. The researcher established ChatGPT accounts for all participants in the experimental group to track their progress and support them in writing their reflections. Students received specific questions to gather the information they needed to integrate into their writing for their final reflection. The instructor developed a standardized questionnaire to guide experimental group students in their use of ChatGPT through a controlled process.

The second evaluation used the same assessment method as the first evaluation to maintain assessment consistency. Students completed their second demonstration class as their final assessment for the last phase, which ran from weeks 12 to 14. Students presented their second assignment plans through lectures, which were evaluated against established assessment criteria.

### Data Analysis

This section covers all numerical and qualitative data collected before and after treatment to measure the treatment's impact precisely. The first part of the statistics evaluates pre-service teachers' performance in their demo classes. Before conducting any tests, it was mandatory to assess data normality to select the appropriate parametric or nonparametric test. Table 1 below presents the results of the Kolmogorov-Smirnov normality test.

**Table 1** *Results of Normality Test for Evaluation 1 and Evaluation of Both Groups*

| | Kolmogorov-Smirnov | | |
|---|---|---|---|
| **Groups** | **Statistic** | **df** | **Sig.** |
| Evaluation 2 | | | |
| Control | .117 | 20 | .200 |
| Experimental | .113 | 20 | .200 |
| Evaluation 1 | | | |
| Control | .112 | 20 | .200 |
| Experimental | .113 | 20 | .200 |

Table 1 shows no significant deviation from normality ($p = 0.200$) for the control group in evaluation 1. Similarly, the evaluation 1 results for the experimental group support the assumption of normality ($p = 0.200$). It may be assumed that the distributions of values in both groups at the first measurement point were normal. Results of evaluation 2 indicated that the p-values for both groups were greater than the significance level of 0.05, and thus, the evaluation 2 scores for both groups were normally distributed. This shows that the normality assumption was consistently satisfied. That means it will

justify using a parametric statistical test in further analysis. Meeting this assumption ensures that these comparisons are statistically meaningful. Therefore, a two-way mixed ANOVA was selected to compare their performance within and between the groups. This enabled simultaneous investigation of between-group differences and changes over time. Table 2 shows the within-group performance from evaluation 1 to evaluation 2.

**Table 2** *Results of Group Performance from Evaluation 1 to Evaluation 2*

| Effect | Value | F | Hypothesis df | Error df | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|---|
| Time | | | | | | |
| Pillai's Trace | .876 | 268.549 | 1.000 | 38.000 | .000 | .876 |
| Wilks' Lambda | .124 | 268.549 | 1.000 | 38.000 | .000 | .876 |
| Hotelling's Trace | 7.067 | 268.549 | 1.000 | 38.000 | .000 | .876 |
| Roy's Largest Root | 7.067 | 268.549 | 1.000 | 38.000 | .000 | .876 |
| Time * groups | | | | | | |
| Pillai's Trace | .630 | 64.718 | 1.000 | 38.000 | .000 | .630 |
| Wilks' Lambda | .370 | 64.718 | 1.000 | 38.000 | .000 | .630 |
| Hotelling's Trace | 1.703 | 64.718 | 1.000 | 38.000 | .000 | .630 |
| Roy's Largest Root | 1.703 | 64.718 | 1.000 | 38.000 | .000 | .630 |

Table 2 shows that all the participants experienced a considerable increase from evaluation 1 to evaluation 2. This suggests that test performance improved over the course of the study regardless of group. This is supported by Pillai's Trace (0.876), Wilks' Lambda (0.124), Hoteling's Trace (7.067), and Roy's Largest Root (7.067), all yielding $F(1,38) = 268.549$, $p < 0.001$. Partial eta squared at 0.876, indicating that 87.6% of the score variance was due to time. Hence, the effect size is very large. Also, the interaction of Time × Groups was significant, as shown by Pillai's Trace (0.630), Wilks' Lambda (0.370), Hotelling's Trace (1.703), and Roy's Largest Root (1.703), all producing $F(1, 38) = 64.718$, with $p < 0.001$. This interaction effect suggests that convergence over time differs in magnitude between the control and experimental groups. Since both groups showed progress over time, it was necessary to compare their performance to assess the impact of treatment; therefore, Table 3 below shows the results of the comparison between the control and experimental groups. The comparison allows for assessing whether the benefits observed differ between those exposed to the intervention.

**Table 3** *Results of Pairwise Comparisons in Both Groups*

| (I) groups | (J) groups | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval for Difference | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| Control | Experimental | –4.225 | 1.515 | .008 | –7.292 | –1.158 |
| Experimental | Control | 4.225 | 1.515 | .008 | 1.158 | 7.292 |

Table 3 shows that the performances between the control and experimental groups are statistically different. This indicates a need for a clear difference at the post-intervention time. The mean difference between the control and experimental groups is –4.225, with an SE of 1.515 and a p-value of 0.008, indicating that the experimental group outperformed the control. The difference between groups is positive, indicating that those who participated in the intervention scored higher on the outcome than those in the control condition. The mean 95% CI for these differences lies between –7.292 and –1.158, excluding zero; thus, the observation is probably true. Because the confidence intervals did not include zero, confidence in the stability of this group difference increased. Reversing the comparison, the values would be 4.225 with the identical p = 0.008 and a CI of [1.158, 7.292]. Symmetry means the effect is the same no matter which way the comparison is made. From the data, it is evident that the improvement in the experimental group far exceeded that of the controls; hence, any intervention accorded to the experimental group further ascertains its effectiveness. Taken together, these results provide direct evidence that this treatment was superior to natural history.

The second part of the statistical evaluation examines the performance of both groups in written reports that reflect their understanding of the lesson plan and its details. Since this was the study's second variable, another normality test was conducted to select suitable tests for comparing progress within and between groups. Table 4 below shows the normality test results for evaluations 1 and 2 of the control and experimental groups.

**Table 4** *Results of the Normality Test for Both Groups in Evaluations 1 and 2*

| | Shapiro-Wilk | | |
|---|---|---|---|
| **Groups** | **Statistic** | **df** | **Sig.** |
| Evaluation 1 | | | |
| Control | .945 | 20 | .297 |
| Experimental | .924 | 20 | .118 |
| Evaluation 2 | | | |
| Control | .965 | 20 | .658 |
| Experimental | .906 | 20 | .054 |

Table 4 for Evaluation 1 and Evaluation 2 for the control and experimental groups, respectively, presents the following statistics: in Evaluation 1, the statistic for the control group is 0.945, df = 20, p = 0.297; for the experimental group, the statistics are 0.924, df = 20, p = 0.118. The p-values in both instances do not indicate a significant deviation from normality. In fact, the baseline score distributions were sufficiently normal in both groups to conduct this analysis. In Evaluation 2, the statistic for the control group is 0.965 (df = 20, p = 0.658), while that for the experimental group is 0.906 (df = 20, p = 0.054). Although the p-value in the experimental group in Evaluation 2 is close to the significance level, the results are still normally distributed. Table 5 below presents participants' performance within their groups using a one-sample test.

Based on Table 5, there is an evident improvement from Evaluation 1 to Evaluation 2 for the control group, as the average score rises from 49.20 in Evaluation 1 to 59.05 in Evaluation 2, an increase of 9.85. However, the increase in the control group suggests that learning may have been a factor, given that they had become familiar with the test. This thus shows a huge improvement between the two assessments. The same trend continues for the experimental group, and the scores are much better between the first and second evaluations, with the mean difference increasing from 49.35 in

**Table 5** *Comparison of Both Groups from Evaluation 1 to Evaluation 2*

| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Evaluation 1 Control | 42.172 | 19 | .000 | 49.20000 | 46.7582 | 51.6418 |
| Evaluation 2 Control | 41.457 | 19 | .000 | 59.05000 | 56.0688 | 62.0312 |
| Evaluation 1 Experiment | 36.769 | 19 | .000 | 49.35000 | 46.5408 | 52.1592 |
| Evaluation 2 Experiment | 43.359 | 19 | .000 | 70.75000 | 67.3348 | 74.1652 |

Evaluation 1 to 70.75 in Evaluation 2, a difference of 21.40, indicating the participants performed considerably better between the two tests. The improvement in performance of the experimental group is much larger than that of the control group, implying that the performance of the experimental group is not simply a result of practice. An Independent Samples Test was conducted to evaluate participants' performance across groups, and the results are shown in Table 6. This allows us to determine whether the differences we observe between groups are significant.

**Table 6** *Performance of Both Groups in the Pretest and Posttest*

| | Levene's T | | t-test for Equality of Means | | | | | 95% Confidence Interval of the Difference | |
|---|---|---|---|---|---|---|---|---|---|
| | F | Sig | t | df | Sig. (2-tailed) | M.D | Std. E | Lower | Upper |
| eva.1 | | | | | | | | | |
| EVA | .618 | .437 | –.317 | 38 | .753 | –.55000 | 1.73581 | –4.06396 | 2.96396 |
| EVNA | | | –.317 | 37.649 | .753 | –.55000 | 1.73581 | –4.06503 | 2.96503 |
| eva.2 | | | | | | | | | |
| EVA | .101 | .753 | –6.159 | 38 | .000 | –12.65000 | 2.05398 | –16.80806 | –8.49194 |
| EVNA | | | –6.159 | 37.945 | .000 | –12.65000 | 2.05398 | –16.80826 | –8.49174 |

As Table 6 shows, in Evaluation 1, Levene's test for equality of variances showed no significant difference in variances ($F = 0.618$, $p = 0.437$); thus, the assumption of equal variances can be accepted. Therefore, using the t-test, Equality of Means provides $t = –0.317$, $df = 38$, and $P = 0.753$; hence, there is no significant difference in performance across the groups. This means that both groups began the study at similar levels of performance. The mean difference is –0.550 with an associated standard error of 1.736, and the 95% confidence interval (–4.064, 2.964) also encompasses zero, further justifying that the groups performed similarly in Evaluation 1. Statistical equivalence indicates that the groups' baselines are homogeneous. On the other hand, the results of Levene's test of Equality of Variances for Evaluation 2 are insignificant ($F = 0.101$, $p = 0.753$), allowing the assumption of equal variances. A t-test for Equality of Means was conducted; results show a significant difference between the groups, $t = –6.159$, $df = 38$, $p < 0.001$. This result, therefore, indicates clear differences between groups in post-intervention performance. The mean difference was –12.650, SD 2.054, 95% confidence interval –16.808, –8.492; since it excludes zero, the experimental group outperforms the control group significantly in Evaluation 2. The size and direction of this difference suggest a strong advantage for the experimental condition. It thus follows from these results that even though the groups performed

comparably in Evaluation 1, by Evaluation 2, the experimental group has shown a significant improvement, bringing into relief the effect of the intervention. These findings give direct statistical support for the effectiveness of the applied treatment.

The third part of this section focuses on the qualitative analysis of students' demo practice across both groups in Evaluations 1 and 2, based on observations by two evaluators. The evaluation criteria showed minimal advancement of the control group, as students made few changes and showed no signs of developing autonomy. The control group students demonstrated understanding of classroom targets, but they did not connect their work to objectives, and their learning remained at a superficial level, requiring external assistance. The students made some improvements in organizing their ideas and sharing their thoughts, but their presentations focused primarily on fundamental content and structural issues, which reduced their overall quality. The control group students were required to follow instructions exactly without showing any critical thinking or flexibility in their assignments. The control group students improved their feedback reception, but their assessment answers remained surface-level because they failed to work directly with the assessment criteria.

The control group achieved limited technological progress because they focused on operational improvements rather than developing new applications or making substantial system changes. Students used fundamental technology tools, but they did not use them to generate new content or enhance their academic performance. The modifications brought small benefits to time management; however, students failed to create clear, in-depth summaries of class material. The group showed slightly greater self-assurance but remained uninvolved and showed no signs of wanting to participate actively. The students demonstrated improved language accuracy in their words, but they continued to make many grammatical errors, which limited their speaking ability. The control group failed to provide clear explanations, as their responses repeated themselves and required substantial support from outside sources. The control group showed little variation in their results between the two assessments, as the differences were not statistically significant and did not indicate meaningful improvement in self-sufficiency or ability.

The experimental group showed significant improvements across all assessments from Assessment 1 to Assessment 2, indicating a strong effect of the intervention. The students demonstrated excellent understanding of the instructional targets, followed the established goals, and took initiative to achieve these objectives through creative methods. The students achieved better work organization and content development, leading to improved presentations that effectively demonstrated concepts through logical, well-organized sequences and greater coherence. The experimental group developed multiple effective strategies for its work, demonstrating the ability to think critically and adapt to new situations when faced with obstacles. The experimental group improved self-assessment skills through peer feedback, resulting in enhanced overall performance. The students in Evaluation 2 applied technology effectively to develop interactive learning approaches, resulting in superior academic performance compared to the control group.

The experimental group showed major progress in starting and finishing their courses. The evaluation process at Evaluation 2 required participants to create both strong opening sections and unified closing parts, which summarized essential information. The participants in the experimental group developed greater self-assurance and interest, which motivated them to engage with their classmates and examiners during classroom activities actively. The experimental group improved their language skills through greater fluency and accuracy, spontaneous communication, and fewer grammatical mistakes. The explanations became more understandable in Evaluation 2, which allowed students to share their thoughts with precision and self-assurance. The experimental group achieved substantial progress across all assessment areas, demonstrating their ability to meet objectives through independent work,

creative thinking, and critical analysis skills that the control group lacked. The experimental group achieved better results than the control group due to this intervention, which proved to be an effective method for improving their academic progress.

## Discussion

The research investigated how ChatGPT affects the teaching demonstrations and written reflections of Omani pre-service teachers. The research results demonstrated that participants in the experimental group achieved better results than those in the control group across all assessment areas, including demonstrations and written reflections, as indicated by quantitative data showing higher average scores and statistically significant results in Evaluation 2. The experimental group showed improved autonomy, creativity, critical thinking, and technological skills, as confirmed by qualitative observations, indicating the intervention's effectiveness.

The experiment group's better results could be attributed to ChatGPT, as participants received immediate, customized feedback, which helped them develop their teaching methods and written assessments to a higher level than standard educational approaches. Additionally, the participants used ChatGPT to explore different teaching methods and receive personalized guidance, enabling them to build their analytical and problem-solving skills. The participants achieved better organization and coherence in their instructional demonstrations and written reflections. Participants improved their technological competencies and enhanced their communication skills, gaining self-assurance when interacting with peers and assessors. The changes in their teaching and learning approach indicate improvement, as ChatGPT provided them with professional development tools that the control group lacked.

The research results from this study align with those of Kusuma et al. (2024), demonstrating that ChatGPT offers pre-service teachers immediate personalized feedback, resulting in measurable improvements in their teaching abilities. In another similar study, Kwack and Im (2023) examined teacher development through ChatGPT implementation, focusing on pre-service teachers' experiences, and found that ChatGPT's flexible design and simple operation in educational settings could make teachers more engaged and motivated. In a similar study, Tunç (2024) examined the effects of ChatGPT on pre-service mathematics teachers' lesson planning and the development of teaching methods and found that participants who received the treatment via ChatGPT produced better results than the control group. The research by Karataş and Yüce (2024) assessed how well ChatGPT enhanced pre-service teachers' teaching approaches and performance and found that teachers who received AI assistance developed their professional skills and teaching competencies more than those who did not receive AI support.

## Conclusion

The research investigates the effects of ChatGPT on pre-service teachers during classroom demonstrations and in their written reflective assignments in Omani educational settings. The research data indicate that participants in the experimental group demonstrated substantial growth in their teaching approaches, written reflections, self-directed learning, and creative and critical thinking abilities, which supports ChatGPT's effectiveness for professional development.

The research results from this study provide significant value to teachers, educational organizations, and curriculum developers. The use of ChatGPT by teachers leads to better teaching demonstrations and reflective practices because it develops their professional development skills of autonomy, creativity, and critical thinking. The implementation of ChatGPT and other AI technologies enables instructors to create new teaching approaches that provide customized support and flexible learning options for

students. Educational institutions should use these research results to revise their teacher-training curricula by incorporating technology-based instructional methods that align with current teaching standards. The implementation of AI technology in educational program development enables curriculum developers to build teaching methods and reflective practices that maintain their value in contemporary education.

There are some limitations in the current study worth mentioning. The research involved 40 pre-service teachers from different cultural and linguistic backgrounds in Oman. The study produced positive results, but researchers should avoid broad conclusions because the data came from a single institution. Future studies should include a larger number of participants from diverse educational environments.

The research design tracks teacher trainees' teaching performance and reflective skills development in one month. The brief duration of the intervention period might not reveal the full impact ChatGPT features can have on teaching methods. Implementing ChatGPT tools throughout an entire school year would likely yield better results in establishing the effectiveness of the tools and in developing reflective teaching skills among educators.

The research evaluates success through pre-service teachers' teaching demonstrations, which they combine with their written reflections. The two essential indicators do not address all aspects of teacher preparation because they fail to measure classroom administration, colleague collaboration, and teacher flexibility during unexpected teaching situations. Research should examine various aspects of teacher training through studies that demonstrate ChatGPT's effectiveness as a professional development tool.

The present study did not explore potential barriers and limitations that pre-service teachers may face in incorporating ChatGPT into their teaching practices. Identifying such barriers, ranging from technical and language barriers to resistance to adopting new technologies, may provide critical insights into designing and developing more user-friendly and accessible AI solutions for educator preparation.

Another limitation was that qualitative measures (for example, interviews, think-aloud protocols, and classroom observations) were not used to investigate learner autonomy, learner confidence, and learner creativity. Future studies can consider using a mixed-methods approach to investigate these constructions.

Finally, this study focuses on pre-service teachers in a formal training context and omits consideration of its potential contribution to in-service teacher development. Relatedly, future research may also consider how practicing teachers can benefit from ChatGPT to enhance their teaching practices, continue professional development, and solve problems in real classrooms. Comparing and contrasting pre-service and in-service educator groups may yield important insights into the varying needs and benefits of incorporating AI at different career levels.

## References

Alger, C. (2006). What went well, what didn't go so well: Growth of reflection in pre-service teachers. *Reflective Practice*, *7*(3), 287–301. https://doi.org/10.1080/14623940600837327

Arefian, M. H., Çomoğlu, I., & Dikilitaş, K. (2024). Understanding Eteachers' experiences of ChatGPT-driven collaborative reflective practice through a community of practice lens. *Innovation in Language Learning and Teaching*, 1–16. https://doi.org/10.1080/17501229.2024.2412769

Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence*, *6*. https://doi.org/10.1016/j.caeai.2024.100226

Azimi, E., Kuusisto, E., Tirri, K., & Hatami, J. (2019). How do student teachers reflect on their practice through practicum courses? A case study from Iran. *Journal of Education for Teaching*, *45*(3), 277–289. https://doi.org/10.1080/09589236.2019.1599511

Bain, J. D., Mills, C., Ballantyne, R., & Packer, J. (2002). Developing reflection on practice through journal writing: Impacts of variations in the focus and level of feedback. *Teachers and Teaching*, *8*(2), 171–196. https://doi.org/10.1080/13540600220127368

Behforouz, B., & Al Ghaithi, A. (2025). Visual language learning: Exploring the effects of artificial intelligence-generated animations on vocabulary and motivational growth. *Technology in Language Teaching & Learning*, *7*(4), 103077. https://doi.org/10.29140/tltl.v7n4.103077

Behforouz, B., Al Ghaithi, A., & Alsaadi, A. (2026). The effect of AI-based instruction on vocabulary retention and learner motivation in higher education. *Technology in Language Teaching & Learning*, *8*, 103599. https://doi.org/10.29140/tltl.2026.103599

Boud, D., Keogh, R., & Walker, D. (2013). *Reflection: Turning experience into learning*. Routledge.

Cai, H., Lu, L., Han, B., Wong, L.-H., & Gu, X. (2024). Exploring pre-service teachers' reflection mediated by an AI-powered teacher dashboard in video-based professional learning: A pilot study. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-024-10442-1

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., … Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning Pedagogical Content Knowledge Teachers' Knowledge for Teaching Science* (pp. 77–94). Springer Nature Singapore. https://doi.org/10.1007/978-981-13-5898-2_2

Deng, J., & Lin, Y. (2022). The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems*, *2*(2), 81–83. https://doi.org/10.54097/fcis.v2i2.4465

ElSayary, A. (2024). An investigation teachers' perceptions of using ChatGPT as a supporting tool for teaching and learning in the digital era. *Journal of Computer Assisted Learning*, *40*(3), 931–945. https://doi.org/10.1111/jcal.12926

Floris, F. D. (2023). Mutual engagement in virtual spaces: Unveiling the role of artificial intelligence in shaping teaching practices. *Modern Journal of Studies in English Language Teaching and Literature*, *5*(2), 99–116. https://doi.org/10.56498/522023557

Garcia Cardenas, D. (2014). Learning networks to enhance reflectivity: Key elements for the design of a reflective network. *International Journal of Educational Technology in Higher Education*, *11*(1), Article 1. https://doi.org/10.7238/rusc.v11i1.1736

Griche, M. S., & Bennis, M. (2026). Generative artificial intelligence integration in foreign language education in higher education: A systematic literature review. *Technology in Language Teaching & Learning*, *8*, 102827. https://doi.org/10.29140/tltl.2026.102827

Hashem, M., Alenezi, F., Althobaiti, A., & Alzahrani, N. (2024). Evaluating ChatGPT-generated lesson plans using an instructional design framework: Implications for AI-assisted teaching. *Research and Practice in Technology Enhanced Learning*. https://doi.org/10.1186/s41039-024-00123-5

Hume, A. (2009). A personal journey: Introducing reflective practice into pre-service teacher education to improve outcomes for students. *Teachers and Curriculum*, *11*, 21–28. https://doi.org/10.15663/tandc.v11i1.42

Karataş, F., & Yüce, E. (2024). AI and the future of teaching: Preservice teachers' reflections on the use of artificial intelligence in open and distributed learning. *International Review of Research in Open and Distributed Learning*, *25*(3), 304–325. https://doi.org/10.19173/irrodl.v25i3.7785

King, P. M., & Kitchener, K. S. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, *39*(1), 5–18. https://doi.org/10.1207/s15326985ep3901_2

Körkkö, M., Kyrö-Ämmälä, O., & Turunen, T. (2016). Professional development through reflection in teacher education. *Teaching and Teacher Education*, *55*, 198–206. https://doi.org/10.1016/j.tate.2016.01.014

Korucu Kis, S., & Kartal, G. (2019). No pain no gain: Reflections on the promises and challenges of embedding reflective practices in large classes. *Reflective Practice*, *20*(5), 637–653. https://doi.org/10.1080/14623943.2019.1651715

Krol, C. A. (1996). Pre-service teacher educators' dialogue journals: What characterizes students' reflective writing and teachers' comments. *Annual Meeting of the Association of Teacher Educators*. https://eric.ed.gov/?id=ED395911

Kusuma, I. P. I., Roni, M., Dewi, K. S., & Mahendrayana, G. (2024). Revealing the potential of ChatGPT for English language teaching: EFL preservice teachers' teaching practicum experience. *Studies in English Language and Education*, *11*(2), 650–670. https://doi.org/10.24815/siele.v11i2.34748

Kwack, H. R., & Im, E. A. (2023). A Study of pre-service teachers' perceptions of the use of ChatGPT in practical gardening training. *Journal of People, Plants, and Environment*, *26*(6), 693–709. https://doi.org/10.11628/ksppe.2023.26.6.693

Leonhard, T., & Rihm, T. (2011). Increasing reflection skills through accompanying events for school internships? Concept and results of a pilot project with teacher training students. *Teacher Education Under Scrutiny*, *4*(2), 240–270. https://doi.org/10.25656/01:14722

Luzano, J. F. P. (2024). Reshaping mathematics instruction via impact of AI chatbots on secondary education pre-service teachers. *International Journal of Studies in Education and Science*, *5*(3), 233–245. https://doi.org/10.46328/ijses.97

McCormack, A. C. (2001). Using reflective practice in teaching dance to pre-service physical education teachers. *European Journal of Physical Education*, *6*(1), 5–15. https://doi.org/10.1080/1740898010060102

Mena-Marcos, J., García-Rodríguez, M.-L., & Tillema, H. (2013). Student teacher reflective writing: What does it reveal? *European Journal of Teacher Education*, *36*(2), 147–163. https://doi.org/10.1080/02619768.2012.713933

Nguyen, Q. D., Fernandez, N., Karsenti, T., & Charlin, B. (2014). What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. *Medical Education*, *48*(12), 1176–1189. https://doi.org/10.1111/medu.12583

Quan, J., Li, R., Wu, W., & Chen, H. (2024). Impact of ChatGPT and prompt on pre-service teachers' technology-enhanced learning design: a mixed method study. *2024 International Symposium on Educational Technology (ISET)*, 91–96.

Quinton, S., & Smallbone, T. (2010). Feeding forward: Using feedback to promote student reflection and learning – a teaching model. *Innovations in Education and Teaching International*, *47*(1), 125–135. https://doi.org/10.1080/14703290903525911

Rogers, R. R. (2001). Reflection in higher education: A concept analysis. *Innovative Higher Education*, *26*(1), 37–57. https://doi.org/10.1023/A:1010986404527

Stiler, G. M., & Philleo, T. (2003). Blogging and blogspots: An alternative format for encouraging reflective practice among pre-service teachers. *Education*, *123*(4).

Suphasri, P., & Chinokul, S. (2021). Reflective practice in teacher education: Issues, challenges, and considerations. *PASAA: Journal of Language Teaching and Learning in Thailand*, *62*, 236–264.

Tan, J. (2013). Dialoguing written reflections to promote self-efficacy in student teachers. *Reflective Practice*, *14*(6), 814–824. https://doi.org/10.1080/14623943.2013.836085

Tunç, M. P. (2024). Examining pre-service mathematics teachers' purposes of using ChatGPT in lesson plan development. *Sakarya University Journal of Education*, *14*(2), Article 2. https://doi.org/10.19126/suje.1476326

Ukrop, M., Švábenský, V., & Nehyba, J. (2019, February 22). Reflective diary for professional development of novice teachers. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. Minneapolis MN USA. https://doi.org/10.1145/3287324.3287448

Ullmann, T. D. (2015). *Automated detection of reflection in texts: A machine learning based approach* [Doctoral dissertation, Open University].

Whipp, J., Wesson, C., & Wiley, T. (1997). Supporting collaborative reflections: Case writing in an Urban PDS. *Teaching Education*, *9*(1), 127–134. https://doi.org/10.1080/1047621970090121

Yu, H. (2024). The application and challenges of ChatGPT in educational transformation: New demands for teachers' roles. *Heliyon*, *10*(2). https://doi.org/10.1016/j.heliyon.2024.e24289

Yu, W. M., & Chiu, M. M. (2019). Influences on the reflection quality of journal writing: An exploratory study. *Reflective Practice*, *20*(5), 584–603. https://doi.org/10.1080/14623943.2019.1651712

Zeichner, K., & Wray, S. (2001). The teaching portfolio in US teacher education programs: What we know and what we need to know. *Teaching and Teacher Education*, *17*(5), 613–621. https://doi.org/10.1016/S0742-051X(01)00017-8