

## PAPER

# Teachers' Perceptions of AI-Based Evaluation and a Suggested Questionnaire

Ejtiead Abu Thabet<sup>1</sup> ,  
Rabab Hamdan<sup>2</sup> ,  
Wajeeh Daher<sup>1</sup> 

<sup>1</sup>An-Najah National University,  
Nablus, Palestine

<sup>2</sup>Palestine Technical  
University, Ramallah,  
Palestine

[rabab.hamdan@ptuk.edu.ps](mailto:rabab.hamdan@ptuk.edu.ps)

## ABSTRACT

Artificial intelligence (AI) tools are being utilized by teachers for various purposes, among which is their use for evaluation. The present study examines teachers' perceptions of this use. It does that by suggesting a relevant questionnaire. Moreover, the research examines the level of teachers' perceptions of this use and whether there are significant differences in these perceptions due to background variables, specifically gender, qualification, specialization, and experience with AI. Two hundred and sixty-four teachers participated in the research. We used R and SPSS 27.00 to analyze the data. The study results indicated that the suggested questionnaire had two components, with good model fit and indices. The gender variables resulted in a significant difference in the component 'assessment with AI' with a small effect size. The qualification variable resulted in a significant difference in the 'Exams prepared with AI' component, with negligible effect size. The AI experience variable made a significant difference only for the "Exams prepared with AI" component, consistently benefiting those with more experience. This effect, however, was negligible. It is recommended that the suggested questionnaire be used by educational researchers in different populations to verify the level of teachers' perceptions of AI-based evaluation across different cultures, in addition to the effect of different background variables on this level.

## KEYWORDS

teachers' perceptions, artificial intelligence (AI), evaluation, assessment, exams, questionnaire

## 1 INTRODUCTION

The teacher and the administrator. Educational researchers are interested in the contribution, the restrictions, and the disadvantages of using artificial intelligence (AI) tools in educational settings. Evaluation practices lie at the core of the teachers' work, where this aspect of AI integration in education has met insufficient attention despite its decisive impact on educational outcomes. The present research is interested in this educational aspect, where it examines teachers' perceptions of the role of AI tools in their evaluation practices. We developed a questionnaire that studies

Thabet, E. A., Hamdan, R., Daher, W. (2026). Teachers' Perceptions of AI-Based Evaluation and a Suggested Questionnaire. *International Journal of Interactive Mobile Technologies (IJIM)*, 20(13), pp. 41–52. <https://doi.org/10.3991/ijim.v20i13.61147>

Article submitted 2026-02-17. Revision uploaded 2026-04-10. Final acceptance 2026-04-16.

© 2026 by the authors of this article. Published under CC-BY.

teachers' perceptions of the role of AI tools as evaluation tools. Doing so, we performed validity and reliability tests for this suggested questionnaire. In addition, we examined the level of teachers' perceptions and the effect of background variables on these perceptions.

In the following, we used the term 'evaluation' to include both 'exams prepared with AI' and 'assessment with AI.'

## 1.1 Artificial intelligence in education

Artificial intelligence is increasingly recognized as a transformative force in educational settings, shaping how knowledge is delivered, personalized, and assessed. The integration of AI tools into educational settings is facilitating personalized learning experiences that cater to the unique needs of individual learners. By analyzing data on students' learning behaviors and preferences, AI systems can provide customized feedback, adaptive learning paths, and targeted educational resources, enhancing overall educational effectiveness and engagement [1].

Researchers have studied the different aspects of AI integration in education, such as the cognitive, the psychological, the affective, and the social aspects of learning and teaching. In the context of the cognitive aspect, researchers examined the accuracy of AI tools in solving content problems, such as [2], who examined the accuracy of ChatGPT in solving Chemistry problems. In the context of the psychological aspect, [3] claimed that digital well-being could be conceptualized in association with generative AI. In the context of the affective aspect, [4] found that working with AI tools, university students enriched their task motivation. In the context of the social aspect, [5] argued that the use of GenAI in training had the potential to substantially boost skills in diagnostic communication. It also could function as an individually tailored, easily scalable resource to complement existing medical education. In the present research, we intend to examine another educational aspect, which is the evaluation related to the teaching practice.

## 1.2 Assessment with AI applications

The role of evaluation and feedback is changing, starting with the increasing trend of online education, particularly after the COVID-19 pandemic. Especially after the introduction of AI applications, the evaluation model has shifted towards assessing student activities that promote their learning outcomes [6]. Moreover, this evaluation helps students become deep learners and prepares them for conceptual understanding in subsequent conceptual evaluations. However, [6] warns grading text-based questions, especially in classes with more than 50 registered students, is considered a daunting and time-consuming task for instructors.

[7] reported that AI-powered educational evaluation tools offered numerous benefits to the teaching practice, including improving the accuracy and efficiency of evaluations. [8] explored how the ChatGPT tool impacts evaluation in engineering education, reporting that the AI tool was capable of generating acceptable responses for many evaluations.

[9] studied teachers' perceptions of AI as a tool to support teaching in Estonian K-12 education. The teachers perceived the AI tools as able to support the identification of challenges related to the evaluation context of teaching. The previous result is in line with [10], who conducted an experimental evaluation analyzing the

performance of AI-generated exams versus human-prepared exams. The results confirmed that content coverage was better in the AI-designed exam (81.3% compared to 72.4%). The study affirmed that AI exams can expand content scope and be more efficient in question generation.

[11] aimed to explore the motivation of university faculty members for redesigning evaluations in the era of Generative AI, besides the challenges they faced, and to develop a conceptual framework to assist them. The research findings stressed the need for professional training for faculty, besides the development of innovative evaluation methods that support human-machine collaboration, as well as evaluations based on critical thinking.

[12] noted that AI approaches, while addressing a range of issues, critically described traditional evaluation practices, so we should discuss whether these approaches pose additional challenges to evaluation practice [12] said that we should consider the following to the use of AI tools in educational evaluation: marginalization of professional expertise, black box accountability, restriction of the pedagogical role of evaluation, evaluation of limited forms of learning, the notion of surveillance pedagogy, and reliance on distributed evaluation models. using self-evaluation, peer evaluation, and continuous evaluation instead of a single final evaluation for the student.

[13] aimed to build a Machine Learning model that distinguishes and selects exam items based on their difficulty and discrimination. The study relied on the Artificial Neural Networks method through binary classification models to categorize the items. They demonstrated the effectiveness of the proposed model, which showed superior performance with an accuracy rate of 96% for item selection.

### 1.3 Research rationale and goals

It can be inferred that AI helps accelerate the process of educational evaluation and provides students with rapid feedback that aids them in understanding and overcoming errors, in addition to reducing personal or mood-based bias when assessing students. Key benefits of using AI in evaluation include alleviating the academic burdens on teachers, enhancing decision-making regarding students, and facilitating the development of improvement plans for education. It is interesting to suggest instruments that examine teachers' perceptions of the use of AI tools in educational evaluation, where the present research attempts to do so. Depending on the literature, we intend to suggest a questionnaire that measures teachers' perceptions of AI tools as evaluation instruments. Doing that, we perform validity and reliability analysis for the new questionnaire, examining the level of teachers' perceptions and analyzing how background variables influenced these perceptions.

## 2 METHODOLOGY

An IRB application was submitted to the IRB board at An-Najah National University, and the research was accepted as fulfilling the requirements of ethical research.

### 2.1 Descriptive statistics of the questionnaire items

Table 1 describes the descriptive statistics of the eighteen questionnaire items.

**Table 1.** Descriptive statistics of the questionnaire items (N = 264)

|  | M    | SD   | Skew  | Kurtosis | SE   |
|--|------|------|-------|----------|------|
| Tests prepared with AI are not a cause for concern.  | 2.97 | 1.29 | -0.41 | -1.28    | 0.08 |
| Tests prepared with AI can replace tests prepared by the human.  | 3.49 | 1.04 | -0.77 | -0.25    | 0.06 |
| AI can make tests that are convenient for the student.   | 3.76 | 0.76 | -1.00 | 1.66     | 0.05 |
| AI tools can create secure tests.  | 3.51 | 0.92 | -1.00 | 0.74     | 0.06 |
| Designing and implementing tests using AI can be burdensome for teachers.  | 3.72 | 0.84 | -0.81 | 0.61     | 0.05 |
| AI algorithms make designing tests more reliable.  | 3.73 | 0.93 | -1.06 | 0.72     | 0.06 |
| AI algorithms help develop tests with feedback to the student.   | 3.67 | 0.84 | -0.95 | 1.19     | 0.05 |
| Teachers can benefit from AI tools in designing texts.   | 3.70 | 0.82 | -0.55 | -0.12    | 0.05 |
| Applying AI in designing tests improves the evaluation outcomes.   | 3.74 | 0.85 | -1.38 | 2.31     | 0.05 |
| Applying AI in providing tests provides teachers with questions appropriate for evaluating skills needed to succeed in the 21st century. | 3.64 | 0.80 | -1.11 | 1.53     | 0.05 |
| ChatGPT can generate acceptable responses to many assessments.   | 3.95 | 0.83 | -1.37 | 2.83     | 0.05 |
| Using AI applications enables teachers to review and grade student assignments more efficiently.   | 3.44 | 0.95 | -0.19 | -1.00    | 0.06 |
| AI tools are capable of supporting assessment to fit new educational innovations.  | 3.47 | 1.02 | -0.43 | -0.74    | 0.06 |
| Teachers should have the knowledge to assess AI decisions.   | 3.09 | 1.04 | 0.08  | -1.00    | 0.06 |
| The use of AI can cause ambiguity in the assessment process.   | 3.48 | 0.97 | -0.54 | -0.41    | 0.06 |
| The use of AI in the assessment process leads to a restriction of the educational role of assessment.                                    | 3.22 | 1.06 | -0.30 | -1.08    | 0.07 |
| AI integration provides support to teachers' assessments.  | 3.20 | 0.93 | -0.67 | -0.61    | 0.06 |
| AI tools integrated into education help predict students' academic performance.  | 3.24 | 0.97 | -0.26 | -0.58    | 0.06 |

## 2.2 The research sample

The research sample consisted of 264 teachers. Table 2 describes their distribution over the different background variables: gender, academic qualification, specialization, and AI experience. Table 2 also describes the values of the background variables. All the participants signed a consent form that details the research description and their right to quit at any time during the research.

**Table 2.** Frequencies of the values of each of the background variables

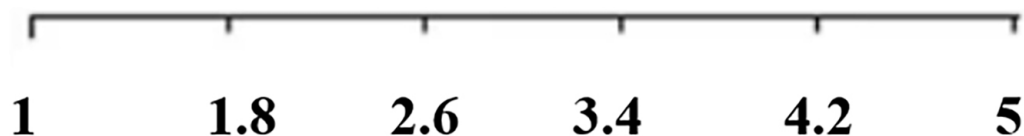
| Variable           | Variable Values | N   |
|--------------------|-----------------|-----|
| Gender             | Male            | 216 |
|                    | Female          | 48  |
| Qualification      | Diploma         | 12  |
|                    | B.A.            | 150 |
|                    | M.A.            | 63  |
|                    | Doctorate       | 39  |
| Specialization     | Humanistic      | 180 |
|                    | Scientific      | 84  |
| Experience with AI | Little          | 81  |
|                    | Medium          | 141 |
|                    | High            | 36  |

### 2.3 Data analysis

Given the sample size was large enough to meet the conditions for parametric tests [14], a violation of the normality assumption would not cause significant issues when calculating the scores for the parametric tests [15]. Thus, we did not test for data normality. To confirm the validity of our analysis, we performed tests for the homogeneity of variance across variables that had more than two values. Levene's tests, across academic qualifications, gave significant results for both sub-scales, indicating non-homogeneity of variance. This test, for 'AI experience,' gave significant results only for the 'Evaluation with AI' sub-scale, indicating non-homogeneity of variance in this component over 'AI experience.' Thus, when non-homogeneity of variance occurred, we computed Welch's F-test, while when homogeneity prevailed, we computed the regular F-test.

We performed the independent sample t-test for background variables that had two values, as in the case of gender and specialization.

To assess the level of teachers' perceptions of AI in educational evaluation, we established a scoring benchmark. We calculated the length of the scoring interval by dividing the total score range (4 units, from the lowest item score of 1 to the highest of 5) by five intervals, resulting in 0.80. This value was used to define the points for various evaluation-perception intervals (see Figure 1). Specifically, we identified the "normal evaluation score" as 2.6 and the "good metacognitive practice score" as 3.4. A score of 4.2 was considered the "very good evaluation perception score." We performed one sample t-test to verify the level of teachers' perceptions of AI-based evaluation.

**Fig. 1.** Intervals related to teachers' perceptions of the evaluation with AI scores

### 3 RESULTS

#### 3.1 Factorizing the questionnaire items

The study initially examined the factorability of 28 items related to teachers' metacognitive skills. After removing items that lacked a correlation of  $\geq 0.4$  with at least one other item, 18 items remained, organized into two factors, one factor with 10 items and the other with eight items. Multiple tests confirmed the factorability of these 18 items: the Kaiser–Meyer–Olkin (KMO) was 0.849, indicating that it is higher than 0.6, the acceptable measure. and Bartlett's test of sphericity was significant ( $p < 0.001$ ). Anti-image correlation diagonals were all above 0.7, and item communalities exceeded 0.4, confirming shared variance. Consequently, a principal factor analysis using Oblimin rotation was conducted on 18 items ( $N = 264$ ). The eigenvalues showed that the first factor (exams prepared with AI) explained 39.962% of the variance of the overall construct, while the second factor (assessment with AI) explained 11.913% of the variance—collectively, the two factors explained 51.875% of the variance. Table 3 displays the resulting factor loadings.

**Table 3.** Factor loading of the items

| Item   | Exams Prepared with AI | Assessment with AI |
|--|------------------------|--------------------|
| Tests prepared with AI are not a cause for concern.  | 0.77                   |                    |
| Tests prepared with AI can replace tests prepared by the human.  | 0.84                   |                    |
| AI can make tests that are convenient for the student.   | 0.67                   |                    |
| AI tools can create secure tests.  | 0.74                   |                    |
| Designing and implementing tests using AI can be burdensome for teachers.  | 0.66                   |                    |
| AI algorithms make designing tests more reliable.  | 0.84                   |                    |
| AI algorithms help develop tests with feedback to the student.   | 0.76                   |                    |
| Teachers can benefit from AI tools in designing texts.   | 0.77                   |                    |
| Applying AI in designing tests improves the evaluation outcomes.   | 0.15                   |                    |
| Applying AI in providing tests provides teachers with questions appropriate for evaluating skills needed to succeed in the 21st century. | 0.62                   |                    |
| ChatGPT can generate acceptable responses to many assessments.   |                        | 0.51               |
| Using AI applications enables teachers to review and grade student assignments more efficiently.   |                        | 0.50               |
| AI tools are capable of supporting assessment to fit new educational innovations.  |                        | 0.50               |
| Teachers should have the knowledge to assess AI decisions.   |                        | 0.74               |
| The use of AI can cause ambiguity in the assessment process.   |                        | 0.55               |
| The use of AI in the assessment process leads to a restriction of the educational role of assessment.                                    |                        | 0.50               |
| AI integration provides support to teachers' assessments.  |                        | 0.50               |
| AI tools integrated into education help predict students' academic performance.  |                        | -0.50              |

Statistics were performed to compute the model fit related to a confirmatory factor analysis (CFA) of the questionnaire. The CMIN/d.f. was 1.63 and thus under 5, indicating a good fit. Moreover, the RMSEA index was 0.04, which succeeded in supporting the model fit of the questionnaire. Bentler's CFI was 0.91 and TLI was 0.92, also indicating that the proposed questionnaire had a good fit.

### 3.2 The level of teachers' perceptions of AI-based evaluation

To assess the significance of the teachers' AI-based evaluation perceptions, a one-sample t-test was performed, comparing the observed score against the pre-defined benchmarks of 2.6 (normal evaluation score) and 3.4 (good evaluation score). See Table 4 for the results.

**Table 4.** Means, standard deviations, and one-sample t-tests for the level of AI-based perception scores

| Component              | M     | SD    | T2.6   | p      | T3.4   | p      |
|------------------------|-------|-------|--------|--------|--------|--------|
| Exams prepared with AI | 3.460 | 0.450 | 30.836 | <0.001 | 2.144  | 0.033  |
| Assessment with AI     | 3.233 | 0.412 | 24.950 | <0.001 | -6.571 | <0.001 |

Table 4 shows that the level of the 'Exams prepared with AI' component is significantly more than the good evaluation score, while the level of the 'Assessment with AI' component is significantly more than the normal evaluation score and less than the good evaluation score.

### 3.3 The difference in teachers' perceptions of AI-based evaluation due to gender

Table 5 shows the results of the independent sample t-test that was conducted to verify whether the differences in teachers' perceptions of AI-based evaluation are significant due to gender.

**Table 5.** Independent sample t-test for the difference in teachers' perceptions of AI-based evaluation due to gender (N = 216 for male teachers, M = 48 for female teachers)

|                        | Gender | M     | SD   | t      | p     | Effect Size |
|------------------------|--------|-------|------|--------|-------|-------------|
| Exams prepared with AI | Male   | 3.437 | .458 | -1.756 | 0.080 | 0.281       |
|                        | Female | 3.563 | .406 |        |       |             |
| Assessment with AI     | Male   | 3.267 | .387 | 2.430  | 0.018 | 0.450       |
|                        | Female | 3.083 | .489 |        |       |             |

Table 5 shows that male teachers have significantly higher scores in 'Evaluation with AI' than female teachers. The effect size is small for gender.

### 3.4 The differences in teachers' perceptions of AI-based evaluation due to qualification

Table 6 shows the results of the ANOVA that was conducted to verify whether the differences in teachers' perceptions of AI-based evaluation are significant due to qualification.

**Table 6.** Results of ANOVA for the differences in teachers' perceptions of AI-based evaluation due to qualification

|                        |           | N   | M     | SD   | F     | p     | Effect Size |
|------------------------|-----------|-----|-------|------|-------|-------|-------------|
| Exams prepared with AI | Diploma   | 12  | 3.278 | .916 | 3.145 | 0.035 | 0.039       |
|                        | B.A.      | 147 | 3.40  | .390 |       |       |             |
|                        | M.A.      | 63  | 3.582 | .484 |       |       |             |
|                        | Doctorate | 39  | 3.539 | .353 |       |       |             |
| Assessment with AI     | Diploma   | 12  | 3.321 | .163 | 1.139 | 0.341 | 0.006       |
|                        | B.A.      | 150 | 3.241 | .398 |       |       |             |
|                        | M.A.      | 63  | 3.184 | .495 |       |       |             |
|                        | Doctorate | 39  | 3.256 | .375 |       |       |             |

Table 6 shows that qualification made a significant difference in the 'Exams prepared with AI' component but not in the 'Evaluation with AI' component of the scale. Saying that, the effect size was negligible. The Games-Howell post hoc exam showed significant differences in the 'Exams prepared with AI' component only between those who have B.A. and those who have M.A.

### 3.5 The differences in teachers' perceptions of AI-based evaluation due to specialization

Table 7 shows the results of the independent sample t-test that was conducted to verify whether the differences in teachers' perceptions of AI-based evaluation are significant due to specialization.

**Table 7.** Independent sample t-test for the difference in teachers' perceptions of AI-based evaluation due to gender (N = 216 for male teachers, M = 48 for female teachers)

|                        |            | Major | M    | SD   | t      | p     | Effect Size |
|------------------------|------------|-------|------|------|--------|-------|-------------|
| Exams prepared with AI | Humanistic | 3.452 | .385 | .566 | -0.355 | 0.723 | 0.054       |
|                        | Scientific | 3.476 | .566 |      |        |       |             |
| Assessment with AI     | Humanistic | 3.234 | .439 | .352 | 0.056  | 0.955 | 0.007       |
|                        | Scientific | 3.231 | .352 |      |        |       |             |

Table 7 shows that the specialization background variable did not result in significant differences in any of the two components of teachers' perceptions of 'Assessment with AI.'

### 3.6 The differences in teachers' perceptions of AI-based evaluation due to AI experience

Table 8 shows the results of the ANOVA that was conducted to verify whether the differences in teachers' perceptions of AI-based evaluation are significant due to 'Experience with AI.'

**Table 8.** Results of ANOVA for the differences in teachers' perceptions of AI-based evaluation due to 'Experience with AI'

|                        |        | N   | M     | SD   | F     | p     | Effect Size |
|------------------------|--------|-----|-------|------|-------|-------|-------------|
| Exams prepared with AI | Little | 81  | 3.325 | .420 | 9.501 | <.000 | 0.069       |
|                        | Medium | 141 | 3.478 | .457 |       |       |             |
|                        | High   | 36  | 3.704 | .406 |       |       |             |
| Assessment with AI     | Little | 81  | 3.318 | .245 | 3.054 | 0.052 | 0.016       |
|                        | Medium | 141 | 3.211 | .436 |       |       |             |
|                        | High   | 36  | 3.206 | .506 |       |       |             |

Table 8 shows that 'Experience with AI' made a significant difference in the 'Exams prepared with AI' component but not in the 'Evaluation with AI' component of the scale. Saying that, the effect size was negligible. The Games-Howell post hoc exam showed significant differences in the 'Exams prepared with AI' component only between each two values of 'Experience with AI' in favor of those who have more experience with artificial intelligence.

## 4 DISCUSSION

Artificial intelligence tools are continually being suggested for use in education [16, 17] and especially for teachers' support [18]. An important issue in teachers' work is evaluation with AI. The present paper came to suggest a questionnaire that measures teachers' perceptions of AI-based evaluation. This questionnaire is composed of two components: Exams prepared with AI and assessment with AI. The questionnaire proved to have a good model fit and indices, making it suitable for use as a tool that measures teachers' perceptions of evaluating educational processes using AI tools.

The results regarding teachers' perceptions of 'AI in Evaluation' showed that the component 'Exams prepared with AI' scored above the "good" benchmark, while the component 'assessment with AI' scored between "normal" and "good." This shows teachers trust AI more in test preparation than in assessing students' work. One possible explanation of this finding is that teachers saw the benefits of preparing exams with AI. [19] reports that teachers' preparation of exams with AI improved students' performance.

The results regarding the effect of gender on teachers' perceptions of 'AI in Evaluation' showed no significant gender difference in "Exams prepared with AI," but male teachers scored significantly higher in "assessment with AI" than female teachers, with a small effect size. These results suggest that male teachers might be slightly more comfortable with AI in assessment. Previous research has found different results regarding teachers' perceptions of working with AI. [20] reported no significant differences between male and female teachers regarding their work with AI. On the other hand, [21] reported that male teachers seem to value GenAI more than female teachers. More research is needed to verify this issue of gender regarding teachers' perceptions of AI.

The results regarding the effect of 'Qualification' on teachers' perceptions of 'AI in Evaluation' showed a significant difference in the 'Exams prepared with AI' component, particularly between B.A. and M.A. holders. Despite this significant difference,

the effect size was negligible. No significant differences in the “Evaluation with AI” component. The significant difference could be explained by the richer academic experience of M.A. holders. This richer academic experience could lead to a deeper and wider understanding of the use of technology in general and AI in particular for educational purposes. Different studies proved the impact of teachers’ qualifications on students’ learning outcomes (ex., Dada [22]), whereas the present research found that this qualification impacts teachers’ perceptions of the use of AI for evaluation.

The results regarding the effect of ‘Specialization’ on teachers’ perceptions of ‘AI in Evaluation’ showed significant differences, with negligible effect size, between humanistic and scientific specializations, in favor of the humanistic specialization only in the ‘assessment with AI’ component. This significant difference could be due to the still some inaccuracy in AI tools’ responses to content problems (ex., Daher et al. [2]) and the relative success of these tools in pedagogic and didactic assignments (ex., Daher and Anabousy [23]).

The results regarding the effect of ‘AI Experience’ on teachers’ perceptions of ‘AI in Evaluation’ showed significant differences, with negligible effect size, only in the ‘Exams prepared with AI’ component. Teachers with more AI experience had significantly higher scores in ‘Exams prepared with AI.’ These results could indicate that familiarity with AI increases confidence in using it for test design. The previous results and claim are in line with [24] and [25], who found that there is a positive correlation between one’s level of experience with AI and the favorability of one’s attitudes toward it.

## 5 CONCLUSIONS, RECOMMENDATIONS, AND LIMITATIONS

Questionnaires are suggested to explore new innovations and practices (e. g., Zhang et al. [26]). The present research intended to suggest a questionnaire that examines teachers’ perceptions of evaluation with AI. It suggested a questionnaire that includes two components and showed how this questionnaire satisfies measures of validity and reliability. These measures make the questionnaire suitable for use by educational researchers to study issues related to evaluation with AI. Using the questionnaire, we studied teachers’ perceptions of evaluation with AI, finding that the levels of these perceptions are normal and good. We can say that the participating teachers embrace AI for efficiency in test preparation, but they are more skeptical about AI’s role in evaluating and giving feedback, possibly due to concerns about fairness and reliability. Thus, there is a need for enriching teachers’ perceptions of evaluation with AI so teachers can benefit from the ability of AI tools in evaluative processes. This enrichment could be done through workshops that discuss evaluation with AI and show how to implement evaluative processes with these tools.

The present research showed that teachers’ perceptions of the two components of ‘evaluation with AI’ due to the demographic variables (gender, qualification, and specialization) were small or negligible, meaning teachers’ perceptions here were relatively consistent across groups. Saying the previous ‘Experience with AI’ background variable was the strongest factor in shaping positive perceptions. The later finding shows the importance of experience with AI. It is expected that teachers, who have little experience with AI, will not trust them with teaching processes, especially in evaluative processes.

The participants in the present research were Palestinians. Future research should study ‘Evaluation with AI’ among other teachers.

## 6 REFERENCES

- [1] A. Nurjanah, I. N. Salsabila, A. Azzahra, R. Rahayu, and N. Marlina, "Artificial Intelligence (AI) usage in today's teaching and learning process: A review," *Syntax Idea*, vol. 6, no. 3, pp. 1517–1523, 2024. <https://doi.org/10.46799/syntax-idea.v6i3.3126>
- [2] W. Daher, H. Diab, and A. Rayan, "Artificial intelligence generative tools and conceptual knowledge in problem solving in chemistry," *Information*, vol. 14, no. 7, p. 409, 2023. <https://doi.org/10.3390/info14070409>
- [3] M. Černý, "Conceptualisation of digital wellbeing associated with generative artificial intelligence from the perspective of university students," *European Journal of Investigation in Health, Psychology and Education*, vol. 15, no. 10, p. 197, 2025. <https://doi.org/10.3390/ejihpe15100197>
- [4] M. Hmoud, H. Swaity, N. Hamad, O. Karram, and W. Daher, "Higher education students' task motivation in the generative artificial intelligence context: The case of ChatGPT," *Information*, vol. 15, no. 1, p. 33, 2024. <https://doi.org/10.3390/info15010033>
- [5] R. X. Suárez-García *et al.*, "DIALOGUE: A generative AI-based pre-post simulation study to enhance diagnostic communication in medical students through virtual type 2 diabetes scenarios," *European Journal of Investigation in Health, Psychology and Education*, vol. 15, no. 8, p. 152, 2025. <https://doi.org/10.3390/ejihpe15080152>
- [6] R. Gao, H. E. Merzdorf, S. Anwar, M. Cynthia Hipwell, and A. Srinivasa, "Automatic assessment of text-based responses in post-secondary education: A systematic review," *Computers & Education: Artificial Intelligence*, vol. 6, p. 100206, 2024. <https://doi.org/10.1016/j.caeai.2024.100206>
- [7] V. J. Owan, K. B. Abang, D. O. Idika, E. O. Etta, and B. A. Basse, "Exploring the potential of artificial intelligence tools in educational measurement and assessment," *EURASIA Journal of Mathematics, Science and Technology Education*, vol. 19, no. 8, p. em2307, 2023. <https://doi.org/10.29333/ejmste/13428>
- [8] S. Nikolic *et al.*, "ChatGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity," *European Journal of Engineering Education*, vol. 48, no. 4, pp. 1–56, 2023. <https://doi.org/10.1080/03043797.2023.2213169>
- [9] I.-A. Chounta, E. Bardone, A. Raudsep, and M. Pedaste, "Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in estonian K-12 education," *International Journal of Artificial Intelligence in Education*, vol. 32, no. 3, pp. 725–755, 2021. <https://doi.org/10.1007/s40593-021-00243-5>
- [10] X. Ma, W. Pan, and X.-N. Yu, "Evaluating AI-generated examination papers in periodontology: A comparative study with human-designed counterparts," *BMC Medical Education*, vol. 25, no. 1, p. 1099, 2025. <https://doi.org/10.1186/s12909-025-07706-6>
- [11] Z. N. Khlaif, W. A. Alkouk, N. Salama, and B. A. Eideh, "Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era," *Education Sciences*, vol. 15, no. 2, p. 174, 2025. <https://doi.org/10.3390/educsci15020174>
- [12] Z. Swiecki *et al.*, "Assessment in the age of artificial intelligence," *Computers and Education: Artificial Intelligence*, vol. 3, no. 3, p. 100075, 2022. <https://doi.org/10.1016/j.caeai.2022.100075>
- [13] N. Hrich, "Artificial intelligence item analysis tool for educational assessment: Case of large-scale competitive exams," *International Journal of Information and Education Technology*, vol. 14, no. 6, pp. 822–827, 2024. <https://doi.org/10.18178/ijiet.2024.14.6.2107>
- [14] A. Ghasemi and S. Zahediasl, "Normality tests for statistical analysis: A guide for non-statisticians," *International Journal of Endocrinology and Metabolism*, vol. 10, no. 2, pp. 486–489, 2012. <https://doi.org/10.5812/ijem.3505>

- [15] J. Pallant, "SPSS survival manual: A step by step guide to data analysis using IBM SPSS," *Australian and New Zealand Journal of Public Health*, vol. 37, no. 6, pp. 597–598, 2020. <https://doi.org/10.1111/1753-6405.12166>
- [16] A. M. Khaddar, A. Dehbi, A. Alali, A. Bakhouyi, and M. Talea, "Emerging technologies in learning: A bibliometric analysis of technology integration and applications," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 19, no. 5, pp. 60–78, 2025. <https://doi.org/10.3991/ijim.v19i05.51267>
- [17] T. Kadhim, H. Al-Ogaili, and A. Mutar, "Integrating artificial intelligence and mobile technologies in Iraqi secondary education," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 19, no. 19, pp. 166–183, 2025. <https://doi.org/10.3991/ijim.v19i19.55791>
- [18] A. D. Samala, X. Zhai, K. Aoki, L. Bojic, and S. Zikic, "An in-depth review of ChatGPT's pros and cons for learning and teaching in education," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 2, pp. 96–117, 2024. <https://doi.org/10.3991/ijim.v18i02.46509>
- [19] M. Dhulipala and A. Bhujade, "Analyzing the impact of AI on exam preparation and student performance in higher education," *Journal of Informatics Education and Research*, vol. 5, no. 2, 2025. <https://doi.org/10.52783/jier.v5i2.2924>
- [20] S. Muhamed and I. Farahana Kamsin, "Teacher's acceptance and intention to use artificial intelligence technology in teaching and learning based on the UTAUT model," *International Journal of Information and Education Technology*, vol. 15, no. 7, pp. 1428–1435, 2025. <https://doi.org/10.18178/ijiet.2025.15.7.2344>
- [21] R. J. Collie and A. J. Martin, "Teachers' motivation and engagement to harness generative AI for teaching and learning: The role of contextual, occupational, and background factors," *Computers and Education. Artificial Intelligence*, vol. 6, p. 100224, 2024. <https://doi.org/10.1016/j.caeai.2024.100224>
- [22] H. O. Yusuf and A. A. Dada, "Impact of teachers' qualification and experience on the performance of students in colleges of education in Kaduna state, Nigeria," *The Online Journal of Quality in Higher Education*, vol. 3, no. 2, pp. 52–78, 2016.
- [23] W. Daher and A. A. Anabousy, "The didactical knowledge of generative artificial intelligence tools: The case of writing mathematics lessons," *Eurasia Journal of Mathematics Science and Technology Education*, vol. 21, no. 9, p. em2691, 2025. <https://doi.org/10.29333/ejmste/16769>
- [24] S. J. Daly, A. Wiewiora, and G. Hearn, "Shifting attitudes and trust in AI: Influences on organizational AI adoption," *Technological Forecasting and Social Change*, vol. 215, no. 215, p. 124108, 2025. <https://doi.org/10.1016/j.techfore.2025.124108>
- [25] S. Lien, E. W. Skogli, J. Y. Abamosa, S. Glomdal, and P. Filkuková, "Experience and attitudes toward the use of artificial intelligence in higher education," *Uniped*, vol. 48, no. 1, pp. 1–14, 2025. <https://doi.org/10.18261/uniped.48.1.1>
- [26] Y. Zhang, J. T. Goddard, and B. A. E. Jakubiec, "Social justice leadership in education: A suggested questionnaire," *Research in Educational Administration & Leadership*, vol. 3, no. 1, pp. 53–86, 2018. <https://doi.org/10.30828/real/2018.1.3>

## 7 AUTHORS

**Ejtiad Abu Thabet** is with the Faculty of Educational Sciences, An-Najah National University, Nablus, Palestine (E-mail: [iabuthabet@najah.edu](mailto:iabuthabet@najah.edu)).

**Rabab Hamdan** is with the Computer Program-Technological Education, Palestine Technical University, Ramallah, Palestine (E-mail: [rabab.hamdan@ptuk.edu.ps](mailto:rabab.hamdan@ptuk.edu.ps)).

**Wajeih Daher** is with An-Najah University, Nablus, Palestine, and Al-Qasemi Academic College of Education, Baqa, Israel (E-mail: [wajeehdaher@najah.edu](mailto:wajeehdaher@najah.edu)).