



RESEARCH ARTICLE

Section: *Literature, Linguistics & Criticism***AI as a language tutor: Jordanian university students' perceptions of AI-generated feedback on their English**Ahmad S. Haider^{1*}¹Department of English Language & Translation, Faculty of Arts & Humanities, Applied Science Private University, Jordan*Correspondence: a_haidar@asu.edu.jo**ABSTRACT**

This study examines Jordanian university students' evaluations of AI-generated feedback on their English and the extent to which these evaluations predict continued-use intention. A questionnaire was completed by 750 students, with analysis of the 25 perception items restricted to the 720 respondents with prior AI experience. The items assessed feedback quality, personalization, usefulness, trust, motivation, and continued-use intention. Exploratory and confirmatory factor analyses supported a six-factor structure. Students rated usefulness ($M = 4.00$) and continued-use intention ($M = 3.98$) more favorably than trust ($M = 3.39$) and personalization ($M = 3.50$). Feedback quality correlated positively with trust ($r = .35$), while personalization correlated with motivation ($r = .41$). Usefulness, trust, and motivation jointly predicted continued-use intention, explaining 14.8% of its variance. Higher self-assessed English proficiency was associated with more favorable evaluations of feedback quality ($\eta^2 = .045$). Formal AI training was associated with slightly higher perceived usefulness ($d = 0.19$), and frequency of use showed a weak positive association with continued-use intention ($p = .14$). Perceived accuracy had little association with usefulness or intention, indicating that students' appreciation of AI feedback did not necessarily reflect confidence in its correctness. These findings support integrating explicit instruction in evaluating AI-generated suggestions into university English programs, retaining teacher guidance and providing additional support for less proficient learners.

KEYWORDS: generative AI, AI-generated feedback, English as a foreign language, trust; personalization, continued-use intention, Jordan

Research Journal in Advanced Humanities

Volume 7, Issue 4, 2026

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted: 29 July 2026

Accepted: 14 September 2026

Published: 14 October 2026

HOW TO CITE

Haider, A. S. (2026). AI as a language tutor: Jordanian university students' perceptions of AI-generated feedback on their English. *Research Journal in Advanced Humanities*, 7(4). <https://doi.org/10.58256/rkkzkn31>

1. Introduction

University students can now obtain feedback on their English by submitting a text to a chatbot. ChatGPT, Gemini, Copilot, and Grammarly identify errors, suggest revisions, explain grammar, and respond to requests for simpler explanations (Kasneci et al., 2023; Kohnke et al., 2023). Earlier educational applications of AI relied on intelligent tutoring and adaptive systems (Chen et al., 2022; Holmes, 2020); generative models have made conversational support accessible through devices that students already use (Fui-Hoon Nah et al., 2023). For learners of English as a foreign language (EFL), this access offers a possible response to familiar difficulties. Research has documented Arab learners' problems with grammar, academic writing, and pronunciation (Atashian & Al-Bahri, 2018; Chaleila & Garra-Alloush, 2019; Kalaldeh, 2016). Teachers' limited time for individual feedback and students' existing use of Google Translate also help explain interest in AI-supported language work (Bin Dahmash, 2020; O'Neill, 2019).

Access to feedback, however, does not ensure that students benefit from it. Learners need to understand a correction, judge its relevance, and decide whether to apply it. These decisions become more difficult when a tool produces fluent but inaccurate language. In a study of Arabic research titles translated into English, Google Translate, Gemini, and ChatGPT all produced semantic and syntactic errors, including mistranslations of polysemous terms (Al-Salman & Haider, 2024). Researchers have also examined the performance of AI tools in translating food idioms (Megdadi et al., 2026) and generating Arabic neologisms for game localization (Hammouri, 2025). Evidence of social bias in AI-generated text adds another reason for scrutiny (Fang et al., 2024; Kotek et al., 2023). The present study examines students' responses to these evaluative demands: whether they trust AI feedback, find it relevant to their needs, and intend to use it again.

Students appear to place different levels of trust in AI depending on what they ask it to do. Gwasmeh et al. (2026) surveyed 2,300 translation students in Jordan and Iraq and documented increasing use for drafting and revising sentences. Respondents trusted AI assistance with grammar but preferred human translators for complex work; those with AI training reported greater trust and ethical awareness. In English–Arabic translation, students also valued the efficiency of ChatGPT (Almahasees et al., 2024). Acceptance of feedback may depend on the instructor's role. In the study by Roe et al. (2024), students and staff were more accepting of AI assessment feedback when instructors added their own comments. Teachers' concerns about standards, ethics, and preparedness further qualify this picture (Almubasher, Abu-Rahme, Abu-Mallouh, et al., 2026; Fadhel et al., 2025; Verano-Tacoronte et al., 2025). Much of this evidence concerns translation or attitudes toward AI in general. It leaves open how students in different disciplines evaluate feedback on their English, and how perceived quality relates to usefulness, trust, personalization, motivation, and continued use.

The study addressed this question through a survey of 750 university students in Jordan. The study drew on the technology acceptance model (Davis, 1989) and the unified theory of acceptance and use of technology (Venkatesh et al., 2003), alongside work on trust in AI (Afroogh et al., 2024; Bach et al., 2024; Choung et al., 2023). The questionnaire measured six perceptions of feedback; their associations with three learner characteristics were also examined. The following research questions guided the study:

RQ1. How do Jordanian university students rate AI feedback on their English in terms of quality, personalization, usefulness, trustworthiness, motivation, and intention to continue using it?

RQ2. What relationships exist among these perceptions, and which predict continued-use intention?

RQ3. How do these perceptions relate to students' self-assessed English proficiency, formal AI training, and frequency of use?

Separate measures of accuracy, usefulness, and continued-use intention allow an examination of whether students value feedback despite doubts about its correctness. Exploratory and confirmatory factor analyses assessed the proposed six-construct structure. Effect sizes were then considered when interpreting the relationships among the measures. This approach distinguishes a favorable judgment of AI feedback from confidence that it is reliable.

2. Literature Review

Intelligent tutoring, adaptive learning, and automated assessment were established educational applications of AI before generative models became available (Chen et al., 2022; Fahimirad & Kotamjani, 2018; Holmes, 2020). With generative models, feedback can be discussed as well as received. A learner can ask why a correction

was made, seek clarification, or request another example (Fui-Hoon Nah et al., 2023; Kohnke et al., 2023). Such interaction may support writing and allow feedback to be tailored to the learner, although the learner still has to evaluate the response. Kasneci et al. (2023) discuss this requirement in relation to inaccurate output and bias, arguing for the development of critical thinking and other competencies needed for effective use.

Digital support for language revision also predates ChatGPT. EFL students have used machine translation to revise writing (Lee, 2020), and second language learners in Saudi Arabia have relied extensively on Google Translate when composing English texts (Bin Dahmash, 2020). Other L2 learners have used online translators together with dictionaries and search engines (O'Neill, 2019). Automatic speech recognition has been studied as a source of pronunciation feedback (Guskaroska, 2019). Chatbots bring several of these functions into an interactive exchange: they can propose a revision and explain it in response to a follow-up question. This makes evaluation of the explanation, as well as the correction, part of the learner's task.

The translation of Arabic research titles illustrates why such judgment is needed. Al-Salman and Haider (2024) recorded fewer errors for Gemini than for ChatGPT or Google Translate, yet all three tools produced errors in meaning or grammar. Reviews of GPT-based translation (Chan & Tang, 2024) and findings on social bias in AI-generated content (Fang et al., 2024; Koteck et al., 2023) provide further grounds for checking output. An incorrect suggestion may be difficult to detect precisely for the learner who needs feedback most. Thus, the availability of a correction does not remove the need for linguistic knowledge to assess it.

In evaluating a correction, students also make a judgment about the tool. Afroogh et al. (2024) link trust to AI adoption and identify accuracy as one technical basis for trustworthiness. The distinction made by Choung et al. (2023) is useful here: human-like trust concerns one form of user response, whereas functionality trust concerns competent and reliable performance. Their findings connect functionality trust with perceived technical robustness. Trust also varies with system design, users, and the socio-ethical setting (Bach et al., 2024). For an English learner, a clear explanation and an apparently correct revision may provide grounds for confidence. The task-specific trust reported by translation students is consistent with this account. The study therefore hypothesized a positive relationship between feedback quality and trust.

Personalization is another established aim of AI in education (Chen et al., 2022; Holmes, 2020). Large language models can respond to a learner's text and follow-up questions and can be prompted to adjust explanations to a particular level (Kasneci et al., 2023; Kohnke et al., 2023). AI-generated media have likewise been proposed for personalized learning and well-being (Pataranutaporn et al., 2021). If students perceive feedback as appropriate to their proficiency and learning needs, it may support their engagement and motivation to practice.

The intended use of a tool is examined in acceptance research through beliefs about what it offers. Perceived usefulness is the strongest belief-level predictor of intention in the technology acceptance model (Davis, 1989). The unified theory of acceptance and use of technology integrates eight earlier models and gives a corresponding role to performance expectancy (Venkatesh et al., 2003). For AI feedback, trust may provide an additional reason to maintain use (Afroogh et al., 2024), while instructor involvement can make the feedback more acceptable (Roe et al., 2024). In the present sample, almost all students already had AI experience. Their intentions therefore concern whether to maintain an existing practice, rather than whether to begin one.

English proficiency is likely to matter when students evaluate AI feedback. Evidence from a different learning context illustrates why learner differences warrant attention. Xu and Jaggars (2014) reported poorer performance in online courses than in face-to-face courses among community college students, with larger differences for some groups, including students with weaker academic preparation. Although that study did not examine AI feedback, its findings caution against expecting technology to benefit all learners equally. An explanation written in English may be accessible to one learner and difficult for another, especially when it relies on grammatical terminology. Proficiency may also affect the specificity of a prompt and the ability to recognize an unsuitable correction.

Preparation may help students address these difficulties. Researchers discussing generative AI in language education recommend explicit instruction in critical and productive use (Kasneci et al., 2023; Kohnke et al., 2023). Calls to incorporate AI ethics and awareness into university language programs in Jordan and Saudi Arabia make a similar argument (Fadhel et al., 2025). Among the translation students discussed earlier, training was associated with greater trust. The study therefore examined whether formal AI training was also associated

with differences in perceived usefulness.

Repeated use gives students further opportunities to assess a tool. Acceptance research treats experience as an influence on the relationship between attitudes and continued use (Venkatesh et al., 2003). A positive relationship between frequency and intention is therefore plausible, although students who already value AI may also choose to use it more often.

Language research has addressed Arabic morphology and Qur'anic semantics (Abu-Rahme et al., 2025; Meqdad et al., 2022) and, more recently, teachers' perspectives on AI standards in education (Almubasher, Abu-Rahme, Abu-Mallouh, et al., 2026), educational software and achievement (Almubasher, Abu-Rahme, Mostafa, et al., 2026), and AI chatbots in relation to self-reliance and adoption (Abu-Rahme et al., 2026). Studying feedback on English extends this work to a practice relevant across university programs. English serves as the medium of instruction in many Jordanian programs, while students enter with different proficiency levels. The emergency transition to online learning during the COVID-19 pandemic also gave students and instructors substantial experience with digital tools (Almahasees et al., 2021).

3. Method

3.1. Research design and hypotheses

The study employed a cross-sectional survey design. Based on the literature reviewed in Section 2, six hypotheses were formulated:

- H1. Perceived quality and accuracy of AI-generated feedback are positively associated with trust and confidence in AI feedback.
- H2. Perceived personalization and adaptability of AI-generated feedback are positively associated with learning motivation and engagement.
- H3. Perceived usefulness, trust and confidence, and learning motivation and engagement each positively predict students' intention to continue using AI-generated feedback.
- H4. Students with different levels of self-assessed English proficiency differ in their perceptions of AI feedback quality and accuracy.
- H5. Students who have received formal AI training differ from those who have not in their perceived usefulness of AI feedback.
- H6. Frequency of AI tool use is positively associated with intention to continue using AI-generated feedback.

3.2. Participants and procedure

The questionnaire was distributed through university learning platforms and student networks. The resulting convenience sample comprised 750 students at Jordanian universities. Thirty respondents (4.0%) had not used AI tools for language learning and completed only the background section. The perception analyses include the other 720 students, who answered every perception item. Their demographic characteristics are presented in Section 4.1.

3.3. Instrument

The questionnaire had two parts. Part I recorded gender, age, field of study, self-assessed English proficiency, frequency of AI use for language learning, the tool used, and formal AI training. Part II contained 25 statements rated from 1 (*strongly disagree*) to 5 (*strongly agree*). Five items measured feedback quality and accuracy (FQ). Each of the remaining constructs had four items: personalization and adaptability (PA), perceived usefulness (PU), trust and confidence (TC), motivation and engagement (ME), and future adoption (FA). Table 3 reproduces the statements.

The statements were developed for this study. The usefulness and future-adoption items drew on perceived usefulness and behavioral intention in acceptance research (Davis, 1989; Venkatesh et al., 2003). The trust items addressed functionality trust, or confidence in the system's competence and reliability (Choung et al., 2023). Items belonging to each construct were presented together; Davis and Venkatesh (1996) found that this arrangement did not meaningfully bias acceptance measures. Translation into Arabic and back-translation into English identified discrepancies that were then resolved. Minor wording adjustments were made after the

questionnaire was piloted with 35 students; these students did not participate in the main survey. The Arabic questionnaire was used for data collection.

3.4. Validity and reliability

The analyses used Python 3 with NumPy, pandas, and SciPy. Skewness values (−0.93 to −0.21) and kurtosis values (−0.81 to 0.35) showed no substantial item-level departures from normality. The Kaiser–Meyer–Olkin (KMO) measure (Cerny & Kaiser, 1977) and Bartlett’s test of sphericity were used to assess suitability for factor analysis. Principal axis factoring with promax rotation examined whether the items reflected six factors. Parallel analysis and the eigenvalue criterion informed factor retention (Laher, 2010). A confirmatory model with six correlated factors was then fitted using maximum likelihood. Fit was assessed against CFI and $TLI \geq .95$, $RMSEA \leq .06$, and $SRMR \leq .08$ (Hu & Bentler, 1999). Cronbach’s alpha and composite reliability (CR) were compared with the .70 threshold (Nunnally, 1978). Average variance extracted (AVE) assessed convergence. Discriminant validity was assessed by comparing each square root of AVE with the interconstruct correlations (Fornell & Larcker, 1981) and examining the heterotrait–monotrait ratio (HTMT) (Henseler et al., 2015). Harman’s single-factor test was included because the measures came from one self-report questionnaire and could be affected by common method variance (Podsakoff et al., 2003).

The KMO value of .854 and significant Bartlett’s test, $\chi^2(300) = 5,927.45$, $p < .001$, supported factor analysis. Eigenvalues exceeded 1 for six factors (4.66 to 1.40). Parallel analysis gave the same retention decision: the sixth eigenvalue, 1.40, was higher than the corresponding random-data threshold of 1.19 at the 95th percentile; the seventh, 0.68, was not. All items loaded on their designated factors after rotation. Pattern coefficients were between .62 and .78, and cross-loadings did not exceed .08. Together, the factors explained 50.0% of item variance.

Fit statistics for the six-factor CFA were $\chi^2(260) = 231.89$, $p = .894$; CFI = 1.000; TLI = 1.006; RMSEA = .000, 90% CI [.000, .007]; and SRMR = .024. All standardized loadings were significant and ranged from .60 to .77. Table 1 reports alpha values of .75 to .84 and CR values of .75 to .84. Reliability therefore exceeded .70 throughout (Nunnally, 1978). The AVE results were less uniform. Usefulness (.566) and motivation (.524) exceeded .50, while feedback quality (.496) and future adoption (.488) were slightly below .50. Personalization (.469) and trust (.431) had lower values. Where AVE was below .50, CR remained above .70, which supports the convergence criterion used here. The trust measure nevertheless requires caution; its lower AVE is considered again in the Conclusion.

For every construct, the square root of AVE (.66 to .75) was greater than its correlations with the other constructs (Table 4). The maximum HTMT value was .52, between personalization and motivation, compared with a conservative threshold of .85. These results supported discriminant validity. The first factor in Harman’s unrotated solution accounted for 18.6% of variance, below 50%. This offers no indication of a dominant general response factor, although the test does not establish that common method variance is absent.

Table 1: Reliability and convergent validity by construct ($n = 720$)

Construct	Items	Standardized loadings	α	CR	AVE
Feedback quality and accuracy (FQ)	5	.66–.75	.83	.83	.50
Personalization and adaptability (PA)	4	.65–.75	.78	.78	.47
Perceived usefulness (PU)	4	.74–.77	.84	.84	.57
Trust and confidence (TC)	4	.60–.72	.75	.75	.43
Motivation and engagement (ME)	4	.71–.74	.81	.81	.52
Future adoption and continued use (FA)	4	.64–.73	.79	.79	.49

Note. α = Cronbach’s alpha; CR = composite reliability; AVE = average variance extracted. CFA fit: $\chi^2(260) = 231.89$, $p = .894$, CFI = 1.000, TLI = 1.006, RMSEA = .000 [.000, .007], SRMR = .024. Item-total correlations were .51 to .69. No item deletion improved its scale’s alpha.

3.5. Data analysis

Each construct score was the mean of its items. Pearson correlations tested H1 and H2; Fisher's z transformation was used to obtain 95% confidence intervals. For H3, a multiple linear regression was estimated, followed by a hierarchical regression with gender, proficiency, training, and frequency entered as controls. Standardized coefficients are accompanied by confidence intervals from 2,000 bootstrap resamples, and variance inflation factors (VIFs) are reported. H4 used one-way ANOVA followed by Tukey HSD comparisons. Welch's ANOVA and the Kruskal–Wallis test checked the results under alternative assumptions. An independent-samples t-test examined H5, with Welch's test used for comparison. For H6, the ordinal frequency measure required Spearman's rank-order correlation. Tests were two-tailed with a significance threshold of .05. Cohen's benchmarks guided effect-size interpretation (Cohen, 1988); correlation strength was also described using Evans (1996).

4. Results

4.1. Participants' profile

Table 2 describes the sample. Females accounted for 50.9% and males for 47.2%; 14 students did not disclose their gender. Most respondents were aged 18–23 (74.7%) and in the first four years of a bachelor's degree (91.1%). English Language and Literature, Translation, and Education students accounted for 64.7% of the sample, so these fields are particularly well represented in the findings. ChatGPT was the most commonly used tool (52.7%). The combined “often” and “very often” responses accounted for approximately half of the sample.

Table 2: *Profile of the sample (N = 750)*

Variable	Category	<i>n</i>	%
Gender	Female	382	50.9
	Male	354	47.2
	Prefer not to say	14	1.9
Age (years)	18–20	248	33.1
	21–23	312	41.6
	24–26	135	18.0
	27 or above	55	7.3
Academic year	First	158	21.1
	Second	185	24.7
	Third	172	22.9
	Fourth	168	22.4
	Fifth or above	42	5.6
	Postgraduate	25	3.3
Field of study	English Language and Literature	215	28.7
	Translation	145	19.3
	Education	125	16.7
	Science, Technology, or Health	115	15.3
	Other Humanities and Social Sciences	110	14.7
	Other	40	5.3
Self-assessed English proficiency	Beginner	65	8.7
	Elementary	115	15.3
	Intermediate	265	35.3
	Upper-intermediate	210	28.0
	Advanced	95	12.7

Frequency of AI use	Never	30	4.0
	Rarely	95	12.7
	Sometimes	245	32.7
	Often	235	31.3
	Very often	145	19.3
Most used AI tool	ChatGPT	395	52.7
	Google Gemini	135	18.0
	Grammarly	95	12.7
	Microsoft Copilot	65	8.7
	Other AI tool	30	4.0
	None (never used AI)	30	4.0
Formal AI training	Yes	285	38.0
	No	465	62.0

Note. Rounded percentages. Formal AI training was reported by 271 (37.6%) of the 720 AI users.

4.2. Students' perceptions of AI feedback (RQ1)

Tables 3 and 4 show the highest ratings for usefulness, $M = 4.00$ ($SD = 0.85$), and continued-use intention, $M = 3.98$ ($SD = 0.80$). Approximately 70% of respondents agreed with the items in these constructs. The mean for motivation was $M = 3.81$, followed by feedback quality at $M = 3.72$. Personalization and trust received lower ratings, with means of $M = 3.50$ and $M = 3.39$, respectively. Students' favorable views were therefore more pronounced for some aspects of feedback than for others.

Agreement was highest for the benefits described in items 12 and 13: task completion (73.1%) and learning efficiency (72.5%). Fewer students endorsed item 17, concerning use without always consulting a teacher (45.6%), or item 15, concerning confidence in applying AI suggestions (46.8%). Between 30.3% and 31.9% of respondents chose a neutral response on the trust items. These responses suggest that perceived practical benefits could coexist with uncertainty about reliability. Agreement with the personalization items was approximately 52%, ranging from 51.8% to 53.2%.

Table 3: Item-level responses to the AI feedback questionnaire ($n = 720$)

Item	Agree %	Neutral %	Disagree %	M	SD
<i>Feedback quality and accuracy (FQ)</i>					
1. AI-generated feedback helps me identify errors in my English.	59.9	26.5	13.6	3.72	1.09
2. AI tools provide accurate corrections to my English language mistakes.	59.4	26.1	14.4	3.69	1.09
3. AI-generated explanations help me understand why my answers are incorrect.	61.7	23.5	14.9	3.74	1.13
4. AI feedback is clear and easy to understand.	61.4	23.5	15.1	3.70	1.10
5. AI tools provide feedback that is relevant to my language learning needs.	61.8	23.5	14.7	3.76	1.11
<i>Personalization and adaptability (PA)</i>					
6. AI tools provide feedback that matches my English proficiency level.	51.8	28.6	19.6	3.50	1.13
7. AI-generated feedback addresses my individual language learning weaknesses.	52.9	26.1	21.0	3.48	1.15
8. AI tools adjust their explanations according to my learning needs.	53.2	27.2	19.6	3.52	1.12

9. AI feedback allows me to learn at my own pace.	52.5	25.8	21.7	3.49	1.17
<i>Perceived usefulness and learning improvement (PU)</i>					
10. AI-generated feedback helps me improve my English writing skills.	68.5	22.1	9.4	3.97	1.05
11. AI feedback helps me develop a better understanding of English grammar.	70.3	19.4	10.3	3.98	1.04
12. Using AI feedback improves my ability to complete English language tasks.	73.1	18.6	8.3	4.01	1.02
13. AI tools make learning English more efficient.	72.5	18.1	9.4	4.05	1.01
<i>Trust and confidence in AI feedback (TC)</i>					
14. I trust the corrections provided by AI language learning tools.	47.8	30.6	21.7	3.42	1.13
15. I feel confident applying AI-generated suggestions to improve my English.	46.8	30.3	22.9	3.38	1.16
16. I consider AI feedback reliable for routine English language learning tasks.	48.2	30.6	21.2	3.41	1.13
17. I feel comfortable using AI feedback without always consulting my English teacher.	45.6	31.9	22.5	3.34	1.15
<i>Learning motivation and engagement (ME)</i>					
18. Receiving immediate AI feedback motivates me to practice English more frequently.	62.5	25.6	11.9	3.77	1.06
19. AI-generated feedback encourages me to participate actively in English learning activities.	64.3	25.1	10.6	3.83	1.05
20. Using AI as a language tutor makes learning English more engaging.	64.9	21.2	13.9	3.82	1.12
21. AI feedback encourages me to take greater responsibility for improving my English.	64.7	22.6	12.6	3.83	1.09
<i>Future adoption and continued use (FA)</i>					
22. I intend to continue using AI tools to receive feedback on my English.	70.0	20.1	9.9	3.98	1.01
23. I plan to use AI-generated feedback more frequently in my future English language learning.	70.0	22.1	7.9	4.00	1.00
24. I would recommend AI-powered language tutoring tools to other English learners.	70.3	19.9	9.9	3.99	1.04
25. I would like AI-generated feedback to be integrated into my English language courses.	70.3	20.3	9.4	3.96	1.02

Note. Agree includes agree and strongly agree responses; Disagree includes disagree and strongly disagree responses. Percentages use the 720 respondents who had previously used AI as the denominator. Rounded percentages may not sum to 100.

Table 4: Construct means, standard deviations, and intercorrelations ($n = 720$)

Construct	M	SD	1	2	3	4	5	6
1. Feedback quality (FQ)	3.72	0.85	(.70)					
2. Personalization (PA)	3.50	0.89	.03	(.69)				
3. Usefulness (PU)	4.00	0.85	.03	.24***	(.75)			
4. Trust (TC)	3.39	0.86	.35***	-.03	.04	(.66)		
5. Motivation (ME)	3.81	0.86	.03	.41***	.26***	.00	(.72)	
6. Future adoption (FA)	3.98	0.80	.06	.14***	.27***	.20***	.26***	(.70)

Note. Parenthesized diagonal entries give the square roots of AVE. The remaining entries are Pearson correlations calculated from construct means. *** $p < .001$; unmarked correlations $p > .09$.

4.3. Relationships among perceptions (RQ2)

H1 and H2. The correlations supported both hypotheses (Table 4). Quality and trust were positively related, $r(718) = .35$, 95% CI [.29, .41], $p < .001$. Personalization was also positively related to motivation and engagement, $r(718) = .41$, 95% CI [.35, .47], $p < .001$. Using Cohen's benchmarks, both effects were medium (Cohen, 1988). Shared variance was approximately 12% for quality and trust and 17% for personalization and motivation.

A different pattern emerged for usefulness and continued-use intention: neither was closely related to feedback quality in Table 4. The result separates students' judgments about accuracy from their judgments about value. It does not assess actual accuracy, and the study included no measure of learning gains. The findings therefore indicate that quality ratings did not closely accompany intention, but do not establish that the feedback improved English or that accuracy has no educational importance.

H3. Usefulness, trust, and motivation jointly predicted future adoption, $F(3, 716) = 41.32$, $p < .001$, $R^2 = .148$, adjusted $R^2 = .144$ (Table 5). Each predictor was significant: usefulness, $\beta = .21$, $p < .001$; trust, $\beta = .19$, $p < .001$; motivation, $\beta = .21$, $p < .001$. H3 was supported. The coefficients were similar, and their bootstrap confidence intervals overlapped substantially; the estimates do not identify a clearly stronger predictor. VIF values were at most 1.07, with a Durbin–Watson statistic of 2.01. The model's f^2 of 0.17 was medium. Approximately 85% of the variance in adoption intention was not accounted for by these measures.

Including quality and personalization did not improve the model (FQ: $\beta = -.02$, $p = .583$; PA: $\beta = .01$, $p = .716$; $R^2 = .148$). Step 1 of the hierarchical analysis entered gender, proficiency, training, and frequency of use. This step explained 2.8% of the variance, $F(4, 715) = 5.15$, $p < .001$, with frequency making the main contribution ($\beta = .16$, $p < .001$). Step 2 entered usefulness, trust, and motivation. Their addition increased R^2 by .142 to .170, $\Delta F(3, 712) = 40.60$, $p < .001$. Their coefficients remained close to those in the original model (PU: $\beta = .22$; TC: $\beta = .18$; ME: $\beta = .20$; all $p < .001$). Training had a small negative coefficient after adjustment ($\beta = -.07$, $p = .045$). Its magnitude and the number of comparisons do not support further interpretation.

Table 5: Multiple regression predicting future adoption intention ($n = 720$)

Predictor	<i>B</i>	<i>SE</i>	β	β 95% CI (bootstrap)	<i>t</i>	<i>p</i>	VIF
(Constant)	1.868	0.192			9.73	< .001	
Perceived usefulness (PU)	0.193	0.034	.21	[.13, .28]	5.72	< .001	1.07
Trust and confidence (TC)	0.179	0.032	.19	[.13, .26]	5.60	< .001	1.00
Motivation and engagement (ME)	0.192	0.033	.21	[.14, .28]	5.82	< .001	1.07

Note. $R^2 = .148$, adjusted $R^2 = .144$, $F(3, 716) = 41.32$, $p < .001$, $f^2 = 0.17$. Bootstrapped confidence intervals are based on 2,000 resamples.

4.4. Differences by learner characteristics (RQ3)

H4. Self-assessed proficiency was associated with differences in quality ratings, $F(4, 715) = 8.33$, $p < .001$, $\eta^2 = .045$, $\omega^2 = .039$ (Table 6), a small-to-medium effect. Levene's test did not indicate unequal variances, $F(4, 715) = 1.10$, $p = .356$. Welch's ANOVA, $F(4, 242.70) = 8.52$, $p < .001$, and the Kruskal–Wallis test, $H(4) = 33.87$, $p < .001$, supported the same conclusion. H4 was supported. The Tukey comparisons separated the upper-intermediate ($M = 3.90$) and advanced ($M = 3.97$) groups from the beginner ($M = 3.45$), elementary ($M = 3.49$), and intermediate ($M = 3.66$) groups. The two higher-proficiency groups rated quality significantly more favorably than the other three.

Table 6: Perceived feedback quality by self-assessed English proficiency ($n = 720$)

Proficiency	<i>n</i>	<i>M</i>	<i>SD</i>	Higher than (Tukey HSD)
Beginner	62	3.45	0.82	—
Elementary	108	3.49	0.86	—

Intermediate	256	3.66	0.87	—
Upper-intermediate	202	3.90	0.78	Beginner ($p = .002$), Elementary ($p < .001$), Intermediate ($p = .021$)
Advanced	92	3.97	0.83	Beginner ($p = .001$), Elementary ($p < .001$), Intermediate ($p = .018$)

Note. ANOVA: $F(4, 715) = 8.33, p < .001, \eta^2 = .045$. No significant pairwise differences were found among the beginner, elementary, and intermediate groups, or between the upper-intermediate and advanced groups (all $p > .38$).

H5. Students with training rated usefulness at $M = 4.10$ ($SD = 0.81$), compared with $M = 3.94$ ($SD = 0.86$) for students without training. The difference was *significant*, $t(718) = 2.46, p = .014, d = 0.19, 95\% \text{ CI } [0.04, 0.34]$. Levene's test gave $F = 1.75, p = .186$. Welch's estimate was similar, $t(593.68) = 2.49, p = .013$. H5 was supported, but the difference amounted to only 0.16 on the five-point scale. Exploratory comparisons showed no significant differences by training for the other five constructs (all $p > .30$). The observed difference in usefulness should therefore be interpreted separately from ratings of trust, accuracy, and intention.

H6. The correlation between frequency and continued-use intention was positive, $\rho(718) = .14, 95\% \text{ CI } [.07, .21], p < .001$ (Table 7). The result supports H6; its magnitude, however, is very weak according to Evans (1996). Mean intention ranged from 3.65 for rare users to 4.13 for very frequent users. The largest separation was between “rarely” and “sometimes.” Trust also correlated weakly with frequency ($\rho = .11, p = .005$). No significant frequency relationships were found for the other constructs.

Table 7: *Future adoption intention by frequency of AI use ($n = 720$)*

Frequency of AI use	<i>n</i>	<i>M</i>	<i>SD</i>
Rarely	95	3.65	0.95
Sometimes	245	3.97	0.79
Often	235	4.04	0.73
Very often	145	4.13	0.77

Note. Spearman's $\rho = .14, 95\% \text{ CI } [.07, .21], p < .001$; Kendall's $\tau_b = .11, p < .001$.

Support for the hypotheses should be read in relation to effect magnitude. Although all six hypotheses met the .05 criterion, H1–H3 had medium effects, whereas H6 had a very weak effect. None of the construct scores differed significantly by gender.

5. Discussion

5.1. Usefulness and qualified trust

The relatively high intention rating does not remove the reservations evident in the trust items. Around 70% of students endorsed benefits for their writing, grammar, and efficiency and expected to keep using AI. Trust was rated lowest, especially on the item addressing use without a teacher's advice. Translation students in the study discussed earlier similarly distinguished grammatical assistance from complex work, for which they preferred human translators. Instructor comments also improved acceptance of AI feedback among students and staff (Roe et al., 2024). The neutral trust responses in the sample are consistent with students maintaining that distinction. Caution about reliability need not indicate rejection of the tool, particularly when errors have been documented in Arabic–English output.

The personalization ratings suggest a further limitation in students' experience. About half considered the feedback appropriate to their level or weaknesses, although personalized learning is an important proposed benefit of AI (Kasneji et al., 2023; Kohnke et al., 2023). Prompts may partly account for this difference between expectation and experience. Asking for corrections supplies a text, but may provide little information about a learner's proficiency or the problems they want to address. A response could then be useful without addressing those particular needs. Students' prompts were not collected, so the explanation remains to be tested.

5.2. *Perceived reliability and perceived value*

Quality and trust formed one relationship in the correlation matrix; personalization and motivation formed another. By comparison, quality showed little association with usefulness or intention. This matters because an intention to continue is sometimes treated as evidence of a favorable overall evaluation. Here, students' ratings distinguish the perceived value of using feedback from confidence in its correctness. A model that links technical quality directly to adoption would not adequately describe this pattern.

Convenience offers one possible explanation. Immediate suggestions may help students complete a task or feel that a text has improved, even when they are unsure about individual corrections. The role of usefulness in acceptance research makes this interpretation plausible (Davis, 1989; Venkatesh et al., 2003), but students were not asked which considerations shaped each judgment. Its educational consequences also remain open. If a learner accepts an inaccurate revision, the revision may reinforce an error. The evaluative competencies discussed by Kasneci et al. (2023) are relevant for that reason. Experience alone should not be assumed to produce them.

5.3. *What the adoption model explains*

The regression identifies contributions from usefulness, trust, and motivation of comparable size. Together, these measures are consistent with cognitive, affective, and motivational accounts of continued use (Afroogh et al., 2024; Venkatesh et al., 2003). The model's explained variance, however, was only $R^2 = .15$. These perceptions are therefore part of the explanation, rather than a sufficient account of continued use. Frequency contributed after adjustment in the hierarchical model. Other candidates include social influence and facilitating conditions, as specified in the unified theory of acceptance and use of technology (Venkatesh et al., 2003). Rules governing AI and the demands of particular courses may also influence students' intentions (Almubasher, Abu-Rahme, Abu-Mallouh, et al., 2026; Fadhel et al., 2025). Including these factors would allow a broader account to be tested.

5.4. *Proficiency, training, and frequency of use*

The proficiency comparison identifies where additional guidance may be needed. Ratings from upper-intermediate and advanced students were higher than those from the three lower-proficiency groups, without a consistent difference at every adjacent level. Understanding grammatical terminology, constructing a specific prompt, and evaluating a proposed correction may each be easier with stronger English. The survey does not distinguish among these explanations. The finding does, however, caution against assuming that students with the greatest linguistic needs will gain the most from unrestricted AI access. Evidence on differences in technology-mediated learning raises a similar concern (Xu & Jaggars, 2014). Teachers could use model prompts and Arabic explanations to help these students interpret feedback and assess whether it is appropriate.

The training difference in usefulness was small ($d = 0.19$). Because students were not assigned to training, prior interest or experience could account for part of the difference. The greater trust reported among trained translation students by Gwasmeh et al. (2026) did not appear here. Neither trust nor accuracy ratings differed with training. This may reflect a focus on operating tools, but there is no record of course content with which to examine that possibility. Work linking digital literacy and self-efficacy to learning behavior, such as that by Prior et al. (2016), indicates that preparation deserves closer study. It does not establish what the training in this sample taught.

The weak relationship between frequency and intention is compatible with the role of experience in acceptance research (Venkatesh et al., 2003). The largest difference was between rare users and other users. Because frequency and intention were measured at the same time, the data cannot establish whether repeated use strengthens intention or whether an existing intention encourages use.

6. Conclusion

Jordanian university students in this sample valued AI-generated feedback on their English and expected to continue using it. Their ratings of trust and personalization were less favorable. Perceived accuracy had little relationship with either usefulness or continued-use intention, and usefulness, trust, and motivation explained less than 15% of the variance in intention. These findings show why continued use and confidence in feedback

need to be considered separately. Neither frequent use nor a favorable usefulness rating demonstrates a gain in English proficiency.

In teaching, students could be asked to justify why they accept or reject an AI suggestion. Comparing a correction with a reference or teacher feedback would provide a basis for accepting or rejecting it. Prompt templates and Arabic explanations could help less proficient students undertake this task with guidance. Training should address these evaluative decisions as well as operation of the tools, in line with Fadhel et al. (2025). The training comparison found no advantage in trust or accuracy ratings. The response to item 17 also indicates reservations about relying on AI without consulting an instructor. This gives a reason to retain teacher involvement, consistent with the greater acceptance of combined feedback reported by Roe et al. (2024).

Several features of the study limit the conclusions that can be drawn. Data collected at a single point cannot establish the direction of an association (Shadish et al., 2002), including that between frequency and intention. The study also lacked learning-gain measures and proficiency tests against which to assess perceived usefulness and self-rated proficiency. Language majors made up a large share of the convenience sample, so the findings may not apply to other student populations in Jordan or elsewhere. Further work on measurement is needed because AVE was furthest below .50 for trust (.43) and personalization (.47). Measurement invariance across proficiency and training groups should also be tested before extending the group comparisons. Finally, the distinction between accuracy–trust and value–engagement emerged from the analysis; it was not specified in advance. Replication is needed to assess the theoretical interpretation offered here.

Students' actual revisions could provide evidence of how their perceptions inform their use of feedback. Comparing their responses to AI and teacher suggestions, together with their explanations for accepting or rejecting each suggestion, would help connect the survey findings to practice. Following students over a semester could show whether trust changes with use. An instructional intervention could examine whether students learn to judge feedback accurately and adjust their confidence accordingly. The adoption model could also be extended to include social influence, facilitating conditions, and institutional policy, testing whether these account for some of the unexplained variance.

References

- Abu-Rahme, M. O., Abu-Mallouh, R., Omeish, F., Hindieh, A., Abu-Loghod, N. A., Khader, J. A., & Joudeh, J. M. (2026). AI chatbots' impact on self-reliance and adoption: A study applied to educational institutions. *International Journal of Data and Network Science*, 10(4), 2123–2134.
- Abu-Rahme, M. O., Al-Hawawsheh, S. H., Meqdad, S. A., & Alhroot, M. (2025). The Morphological Reversion of Intensive Patterns in the Poems of Al-Mutanabbi in Praise of Sayf Al-Dawla Al-Hamdani. *Theory and Practice in Language Studies*, 15(4), 1226–1232. <https://doi.org/10.17507/tpls.1504.21>
- Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications*, 11(1), 1–30. <https://doi.org/10.1057/s41599-024-04044-8>
- Al-Salman, S., & Haider, A. S. (2024). Assessing the accuracy of MT and AI tools in translating humanities or social sciences Arabic research titles into English: Evidence from Google Translate, Gemini, and ChatGPT. *International Journal of Data and Network Science*, 8(4), 2483–2498. <https://doi.org/10.5267/j.ijdns.2024.5.009>
- Almahasees, Z., Abu-Rayyash, H. K., Olimat, S. N., & Mahadin, D. K. (2024). An Analytical Study on Student Perceptions of Using ChatGPT in Language Translation. *Language Value*, 17(2), 1–23. <https://doi.org/10.6035/languagev.7925>
- Almahasees, Z., Mohsen, K., & Omer, M. (2021). Faculty's and Students' Perceptions of Online Learning During COVID-19. *Frontiers in Education*, 6, 638470. <https://doi.org/10.3389/educ.2021.638470>
- Almubasher, Z. M., Abu-Rahme, M. O., Abu-Mallouh, R., Al-Hawawsheh, S. H., Obeid, M. A., Dabash, A. M. K., & Al-Salman, A. M. (2026). Language teachers' perspectives on artificial intelligence in education: administrative and pedagogical standards in Jordan. *Research Journal in Advanced Humanities*, 7(2), 1–15. <https://doi.org/10.58256/q6wf3862>
- Almubasher, Z. M., Abu-Rahme, M. O., Mostafa, F. N. B., Abu-Mallouh, R. M., & Al-Hawawsheh, S. H. (2026). The impact of educational software on student learning and academic achievement: Evidence from Jordan. *Research Journal in Advanced Humanities*, 7(3), 1–16. <https://doi.org/10.58256/60kvj878>
- Atashian, S., & Al-Bahri, K. (2018). Towards Arab Students' Grammatical Errors in Academic Writing & their Perceptions. *Arab World English Journal*, Proceedings of 1st MEC TESOL Conference 2018, 140–145. <https://doi.org/10.24093/awej/MEC1.10>
- Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2024). A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *International Journal of Human–Computer Interaction*, 40(5), 1251–1266. <https://doi.org/10.1080/10447318.2022.2138826>
- Bin Dahmash, N. (2020). 'I Can't Live Without Google Translate': A Close Look at the Use of Google Translate App by Second Language Learners in Saudi Arabia. *Arab World English Journal*, 11(3), 226–240. <https://doi.org/10.24093/awej/vol11no3.14>
- Cerny, B. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, 12(1), 43–47. https://doi.org/10.1207/s15327906mbr1201_3
- Chaleila, W., & Garra-Alloush, I. (2019). The most frequent errors in academic writing: A case of EFL undergraduate Arab students in Israel. *English Language Teaching*, 12(7), 120–138. <https://doi.org/10.5539/elt.v12n7p120>
- Chan, V., & Tang, W. K.-W. (2024). GPT for translation: A systematic literature review. *SN Computer Science*, 5(8), 986. <https://doi.org/10.1007/s42979-024-03340-z>
- Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education. *Educational Technology & Society*, 25(1), 28–47. https://www.j-ets.net/collection/published-issues/25_1
- Choung, H., David, P., & Ross, A. (2023). Trust and ethics in AI. *AI & Society*, 38(2), 733–745. <https://doi.org/10.1007/s00146-022-01473-4>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum. <https://doi.org/10.4324/9780203771587>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology

- acceptance model: three experiments. *International Journal of Human-Computer Studies*, 45(1), 19–45. <https://doi.org/10.1006/ijhc.1996.0040>
- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co. <https://search.worldcat.org/title/32465263>
- Fadhel, A. M., Alyami, N., & Almahasees, Z. (2025). Integrating AI Translation Ethics and Sociolinguistic Awareness into BA Programs at Saudi and Jordanian Universities. *Forum for Linguistic Studies*, 7(6), 161–174. <https://doi.org/10.30564/fls.v7i6.9409>
- Fahimirad, M., & Kotamjani, S. S. (2018). A review on application of artificial intelligence in teaching and learning in educational contexts. *International Journal of Learning and Development*, 8(4), 106–118. <https://doi.org/10.5296/ijld.v8i4.14057>
- Fang, X., Che, S., Mao, M., Zhang, H., Zhao, M., & Zhao, X. (2024). Bias of AI-generated content: an examination of news produced by large language models. *Scientific Reports*, 14(1), 5224. <https://doi.org/10.1038/s41598-024-55686-2>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.1177/002224378101800104>
- Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277–304. <https://doi.org/10.1080/15228053.2023.2233814>
- Guskaroska, A. (2019). *ASR as a tool for providing feedback for vowel pronunciation practice* [Master's thesis, Iowa State University]. <https://lib.dr.iastate.edu/etd/17020>
- Gwasmeh, M., Riziq, D. K., Al-Abbas, L. S., Almubasher, Z. M., Haider, A. S., Sahari, Y., & Al-Dulaimi, W. M. (2026). AI in translation classrooms: examining student practices in Jordanian and Iraqi universities. *Frontiers in Education*, 11, 1828251. <https://doi.org/10.3389/feduc.2026.1828251>
- Hammouri, D. M. (2025). The role of artificial intelligence in generating Arabic neologisms: A study of gaming localization. *Research Journal in Advanced Humanities*, 6(4), 1–15. <https://doi.org/10.58256/daewb786>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Holmes, W. (2020). Artificial intelligence in education. In A. Tatnall (Ed.), *Encyclopedia of education and information technologies* (pp. 88–103). Springer. https://doi.org/10.1007/978-3-030-10576-1_107
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kalaldeh, R. (2016). English pronunciation errors by Jordanian university students. *Arab World English Journal*, 7(2), 394–416. <https://doi.org/10.2139/ssrn.2814830>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Kotek, H., Dockum, R., & Sun, D. (2023). Gender bias and stereotypes in large language models. In *Proceedings of the ACM Collective Intelligence Conference* (pp. 12–24). <https://doi.org/10.1145/3582269.3615599>
- Laher, S. (2010). Using exploratory factor analysis in personality research: Best-practice recommendations. *SA Journal of Industrial Psychology*, 36(1), 1–7. <https://doi.org/10.4102/sajip.v36i1.873>
- Lee, S.-M. (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, 33(3), 157–175. <https://doi.org/10.1080/09588221.2018.1553186>
- Megdadi, S., Almahasees, Z., Al-Harashsheh, A., Al Rousan, R., Al-Taher, M. A., & Husienat, I. (2026). Translating food idioms across the English-Arabic language divide: A comparative study of human

- translators, ChatGPT, and Gemini. *Research Journal in Advanced Humanities*, 7(2), 1–15. <https://doi.org/10.58256/pf7v7z49>
- Meqdad, S. A., Abu-Rahme, M. O., Al-Hawawsheh, S. H., & Al-Taher, M. A. (2022). The Semantic Connotation of Contradictory Contrast in the Noble Qur'an, Represented in Serr (Concealment) and Alan (announcement). *Theory and Practice in Language Studies*, 12(7), 1410–1418. <https://doi.org/10.17507/tpls.1207.22>
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill. <https://search.worldcat.org/title/3167590>
- O'Neill, E. M. (2019). Online translator, dictionary, and search engine use among L2 students. *CALL-EJ: Computer-Assisted Language Learning–Electronic Journal*, 20(1), 154–177. <https://callej.org/index.php/journal/article/view/265>
- Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., & Sra, M. (2021). AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence*, 3(12), 1013–1022. <https://doi.org/10.1038/s42256-021-00417-9>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Prior, D. D., Mazanov, J., Meacheam, D., Heaslip, G., & Hanson, J. (2016). Attitude, digital literacy and self efficacy: Flow-on effects for online learning behavior. *The Internet and Higher Education*, 29, 91–97. <https://doi.org/10.1016/j.iheduc.2016.01.001>
- Roe, J., Perkins, M., & Ruelle, D. (2024). Is GenAI the future of feedback? Understanding student and staff perspectives on AI in assessment. *Intelligent Technologies in Education*. <https://doi.org/10.70770/rzzz6y35>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin. <https://search.worldcat.org/search?q=bn:0395615569>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Verano-Tacoronte, D., Bolívar-Cruz, A., & Sosa-Cabrera, S. (2025). Are university teachers ready for generative artificial intelligence? Unpacking faculty anxiety in the ChatGPT era. *Education and Information Technologies*, 30(14), 20495–20522. <https://doi.org/10.1007/s10639-025-13585-7>
- Xu, D., & Jaggars, S. S. (2014). Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. *The Journal of Higher Education*, 85(5), 633–659. <https://doi.org/10.1080/00221546.2014.11777343>