Leveraging AI Tools to Improve Saudi EFL Students' Academic Writing Skills: Instructor Perceptions

Hammad Alshammari[1,*]

[1]English Department, College of Arts, Jouf University, Sakaka, Saudi Arabia

*Correspondence: English Department, College of Arts, Jouf University, Sakaka, Saudi Arabia. Tel: 966-507-145-582. E-mail: h.alshammari@ju.edu.sa

Received: November 4, 2025 Accepted: January 19, 2026 Online Published: February 7, 2026

doi:10.5430/jct.v15n1p267 URL: https://doi.org/10.5430/jct.v15n1p267

Abstract

This study explores instructors' perceptions about integrating AI-driven tools into the curriculum in order to improve the academic writing skills of English as a foreign language (EFL) students in Saudi Arabia. A mixed-methods approach was adopted, implementing a 16-item online survey based on four primary themes: learner familiarity with AI tools, teacher familiarity with AI tools, learning materials, and assessment practices. The results showed that an instructor's years of experience teaching played a major role in their acceptance of AI-driven tools. Instructors with more experience reported a more positive perception of the impact of AI on Saudi EFL learners' writing. Based on the findings, this study recommends providing interactive professional training for both instructors and learners, ensuring availability of the proper tools, and interactive evaluation.

Keywords: AI-driven tool, Saudi EFL learner, technology integration, teacher perceptions, writing skills

1. Introduction

Despite many years of robust government funding for English as a foreign language (EFL) education in Saudi Arabia, learning outcomes routinely fall below expectations, including writing skills (Alkodimi & Al-Ahdal, 2021; Alshammari, 2022; Khan et al., 2020). Few studies have investigating how AI-driven tools could improve the writing skills of Saudi EFL students or how years of experience could affect instructors' perceptions of this technology (McMullen, 2009; Mohammed & Ali, 2021). The present study has sought to address this gap.

1.1 Related Areas

Chen et al. (2024) conducted a systematic review of 30 SSCI-indexed articles on immersive technology in writing education from Web of Science, focusing on theoretical foundations, sampling, technology type, methods, and findings. Technology integration in writing education was mostly perceived positively. Aldosemani et al. (2023) reviewed 16 studies on automated writing evaluation, offering critical insights on feedback quality, consistency, and usefulness. They concluded that this type of evaluation could enhance EFL writing performance.

1.2 Digital Multimodal Composition

Abdelhalim (2024) explored the perceived impact of collaborative digital multimodal composition on the writing strategies of 50 Saudi high school EFL students, who were monitored during a nine-week intervention. This mixed-methods study, involving a self-regulated writing strategies questionnaire and interview, found largely positive perceptions.

Following similar procedures, Khan and Kumar (2023) investigated the impact of metacognitive writing strategies and online instruction on 80 intermediate-level EFL college students from different writing levels in Saudi Arabia. Data were collected through a metacognitive survey and writing task. Metacognitive strategies had a positive impact on writing consistency and cohesion.

Daweli (2018) examined how exposing Saudi advanced EFL students to Google Docs and online peer review influenced academic writing. Data were collected via practice with Google Docs, an online survey, and an interview. Hierarchical power, learners' prior beliefs, and experience were found to influence learner feedback. Alsahil (2024) studied learners' perceptions about online collaborative writing through Google Docs as well. Participants perceived

Google Docs to have academic and technical benefits but no social benefit.

1.3 AI-Driven Tools

Kwon et al. (2023) explored 75 Korean second language (L2) learners' perceptions about using a chatbot in L2 writing. A control group used traditional writing practices, while an experimental group employed Google's Dialogflow machine-learning tool for 15 weeks. The experimental group had significantly higher post-test scores, and those participants claimed that chatbots could improve writing skills. Lee et al. (2024) investigated the perceptions of Korean EFL learners toward the AI-driven tools such as Google Translate, Naver Papago, and Grammarly, collecting data through an online survey, interview, and focus group. The findings suggested that such tools could improve EFL writing skills.

Mohammed and Ali (2021) surveyed 80 Saudi EFL learners about their attitudes toward informal digital vocabulary acquisition through a self-reported questionnaire. They found a strong correlation between learners' positive attitudes and beneficial writing practices.

McMullen (2009) investigated the impact of major and gender on learners' use of language learning strategies. The 165 participants (94 female, 71 male) were drawn from three Saudi universities, and data were collected through the self-reported Strategy Inventory for Language Learning questionnaire. Students majoring in computer science used more language learning strategies than those majoring in management information systems. Such strategies were also slightly more common among female students than male students.

1.4 Summary

While previous studies have questioned the impact of AI-driven tools on writing practices (e.g., Abdelhalim, 2024; Aldosemani et al., 2023; Alhaider, 2023; Almusharraf & Bailey, 2023; Chen et al., 2024; Kwon et al., 2023; Lee et al., 2024; Muftah, 2023; Riswanto et al., 2023; Sherkuziyeva et al., 2023), most explored learner perceptions with little focus on actual practices (e.g., Aldosemani et al., 2023; Alhaider, 2023; Lee et al., 2024). Others focused on the positive impact of collaborative learning on EFL learners' writing skills (e.g., Abdelhalim, 2024; Daweli, 2018; Khan & Kumar, 2023). A small number of studies in Saudi Arabia (e.g., McMullen, 2009; Mohammed & Ali, 2021) and South Korea (e.g., Kwon et al., 2023; Lee et al., 2024) have suggested that such technology could improve EFL learners' writing practices.

2. Methodology

2.1 Research Questions

This mixed-methods study sought to answer two research questions:

1) How do Saudi instructors with different years of experience perceive the impact of AI-driven tools on the writing skills of Saudi EFL students?

2) What suggestions do these instructors have for improving Saudi EFL students' writing skills with AI-driven tools?

2.2 Sample

The participants consisted of 30 EFL faculty members with three levels of experience teaching EFL at the university level: low (less than five years), moderate (5–10 years), and high (more than 10 years), as shown in Table 1.

Table 1. Distribution of Participants

Experience level	*N*	Years of experience	Means' calculation %	Gender	
				Male	Female
High	11	> 10	12	60%	40%
Moderate	11	5–10	6	55%	45%
Low	8	< 5	3.5	50%	50%

2.3 Data Collection

The quantitative data were collected through a survey of 15 Likert-scale items (1 = strongly agree, 2 = agree, 3 = not sure/neutral, 4 = disagree, 5 = strongly disagree) and one open-ended question, internally constructed to target the following themes: learner familiarity with AI tools (Items 1–5), teacher familiarity with AI tools (Items 6–10), learning materials (Items 11–13), and assessment practices (Items 14–15). It was reviewed by two associate

professors in applied linguistics. Based on their suggestions, a few changes were made and one item was removed from the initial version. The participants were contacted about the study directly by the researcher or indirectly through colleagues. Each signed an electronic consent form assuring the anonymity and confidentiality of the data.

The qualitative data consisted of answers to an open-ended survey question and were recorded via the Voice Memos application.

2.4 Data Analysis

Quantitative data were analyzed through SPSS (Version 29.0.2.0). Reliability was determined using Cronbach's alpha, statistical comparisons were made using one-way ANOVA, and post hoc Tukey HSD was employed to assess the significance ratio between different teaching experience levels. The qualitative data were transcribed and analyzed for relevant themes and then assessed by two experienced external referees to increase validity.

3. Results

3.1 Quantitative Results

The internal reliability for each item in the survey is shown in Table 2. Only Item 7 (Cronbach's alpha = .968) was slightly above the cumulative reliability level (Cronbach's alpha = .966). Thus, the 15 Likert-scale survey items showed a high level of internal consistency.

Table 2. Cronbach's Alpha Results

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
1	29.23	148.254	.911	.931	.962
2	29.20	153.683	.866	.863	.963
3	29.23	152.185	.880	.916	.962
4	29.27	154.823	.838	.869	.963
5	29.33	151.678	.855	.896	.963
6	29.57	154.185	.872	.891	.963
7	29.47	166.189	.568	.671	.968
8	29.37	158.723	.767	.820	.965
9	29.33	158.230	.775	.723	.964
10	29.43	160.047	.685	.850	.966
11	29.47	162.464	.746	.801	.965
12	29.37	159.689	.727	.604	.965
13	29.20	157.890	.754	.824	.965
14	29.27	160.133	.762	.664	.965
15	29.33	151.057	.924	.916	.962

Since the internal reliability showed an acceptable level of consistency (i.e., .966) above .00 and below 1, a one-way ANOVA was performed, revealing a significant difference (at p <.005) between groups in responding to Items 1–15, as shown in Table 3.

Table 3. One-way ANOVA Results

Item		Sum of Squares	*df*	Mean Square	*F*	Sig.
1	Between Groups	36.982	2	18.491	36.130	< .001
	Within Groups	13.818	27	.512		
	Total	50.800	29			
2	Between Groups	24.037	2	12.019	24.345	< .001
	Within Groups	13.330	27	.494		
	Total	37.367	29			
3	Between Groups	31.743	2	15.872	47.316	< .001
	Within Groups	9.057	27	.335		
	Total	40.800	29			
4	Between Groups	23.383	2	11.691	24.692	< .001
	Within Groups	12.784	27	.473		
	Total	36.167	29			
5	Between Groups	30.109	2	15.055	27.858	< .001
	Within Groups	14.591	27	.540		
	Total	44.700	29			
6	Between Groups	24.864	2	12.432	31.660	< .001
	Within Groups	10.602	27	.393		
	Total	35.467	29			
7	Between Groups	9.728	2	4.864	11.685	< .001
	Within Groups	11.239	27	.416		
	Total	20.967	29			
8	Between Groups	18.719	2	9.359	22.669	< .001
	Within Groups	11.148	27	.413		
	Total	29.867	29			
9	Between Groups	14.836	2	7.418	12.626	< .001
	Within Groups	15.864	27	.588		
	Total	30.700	29			
10	Between Groups	17.227	2	8.614	15.743	< .001
	Within Groups	14.773	27	.547		
	Total	32.000	29			
11	Between Groups	11.694	2	5.847	17.025	< .001
	Within Groups	9.273	27	.343		
	Total	20.967	29			
12	Between Groups	16.003	2	8.002	15.583	< .001
	Within Groups	13.864	27	.513		
	Total	29.867	29			
13	Between Groups	19.730	2	9.865	19.533	< .001
	Within Groups	13.636	27	.505		
	Total	33.367	29			
14	Between Groups	16.110	2	8.055	21.625	< .001
	Within Groups	10.057	27	.372		
	Total	26.167	29			
15	Between Groups	30.836	2	15.418	42.205	< .001
	Within Groups	9.864	27	.365		
	Total	40.700	29			

Since the one-way ANOVA showed a significant difference between participants based on years of experience, a multiple comparison post hoc Tukey's HSD test was employed to determine which between-group differences were significant. The results are presented in Table 4.

Table 4. Tukey's HSD Results

Dependent Variable	(I) Proficiency level	(J) Proficiency level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Item 1	High	Moderate	-.545	.305	.193	-1.30	.21
		Low	-2.727*	.332	< .001	-3.55	-1.90
	Moderate	High	.545	.305	.193	-.21	1.30
		Low	-2.182*	.332	< .001	-3.01	-1.36
	Low	High	2.727*	.332	< .001	1.90	3.55
		Moderate	2.182*	.332	< .001	1.36	3.01
Item 2	High	Moderate	-.727	.300	.056	-1.47	.02
		Low	-2.261*	.326	< .001	-3.07	-1.45
	Moderate	High	.727	.300	.056	-.02	1.47
		Low	-1.534*	.326	< .001	-2.34	-.72
	Low	High	2.261*	.326	< .001	1.45	3.07
		Moderate	1.534*	.326	< .001	.72	2.34
Item 3	High	Moderate	-.455	.247	.176	-1.07	.16
		Low	-2.511*	.269	< .001	-3.18	-1.84
	Moderate	High	.455	.247	.176	-.16	1.07
		Low	-2.057*	.269	< .001	-2.72	-1.39
	Low	High	2.511*	.269	< .001	1.84	3.18
		Moderate	2.057*	.269	< .001	1.39	2.72
Item 4	High	Moderate	.182	.293	.811	-.55	.91
		Low	-1.898*	.320	< .001	-2.69	-1.10
	Moderate	High	-.182	.293	.811	-.91	.55
		Low	-2.080*	.320	< .001	-2.87	-1.29
	Low	High	1.898*	.320	< .001	1.10	2.69
		Moderate	2.080*	.320	< .001	1.29	2.87
Item 5	High	Moderate	-.273	.313	.663	-1.05	.50
		Low	-2.386*	.342	< .001	-3.23	-1.54
	Moderate	High	.273	.313	.663	-.50	1.05
		Low	-2.114*	.342	< .001	-2.96	-1.27
	Low	High	2.386*	.342	< .001	1.54	3.23
		Moderate	2.114*	.342	< .001	1.27	2.96
Item 6	High	Moderate	-.091	.267	.938	-.75	.57
		Low	-2.102*	.291	< .001	-2.82	-1.38
	Moderate	High	.091	.267	.938	-.57	.75
		Low	-2.011*	.291	< .001	-2.73	-1.29
	Low	High	2.102*	.291	< .001	1.38	2.82
		Moderate	2.011*	.291	< .001	1.29	2.73

Table 4. Tukey's HSD Results(continued)

Dependent Variable	(I) Proficiency level	(J) Proficiency level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Item 7	High	Moderate	-.364	.275	.396	-1.05	.32
		Low	-1.420*	.300	< .001	-2.16	-.68
	Moderate	High	.364	.275	.396	-.32	1.05
		Low	-1.057*	.300	.004	-1.80	-.31
	Low	High	1.420*	.300	< .001	.68	2.16
		Moderate	1.057*	.300	.004	.31	1.80
Item 8	High	Moderate	-.091	.274	.941	-.77	.59
		Low	-1.830*	.299	< .001	-2.57	-1.09
	Moderate	High	.091	.274	.941	-.59	.77
		Low	-1.739*	.299	< .001	-2.48	-1.00
	Low	High	1.830*	.299	< .001	1.09	2.57
		Moderate	1.739*	.299	< .001	1.00	2.48
Item 9	High	Moderate	-.273	.327	.685	-1.08	.54
		Low	-1.705*	.356	< .001	-2.59	-.82
	Moderate	High	.273	.327	.685	-.54	1.08
		Low	-1.432*	.356	.001	-2.31	-.55
	Low	High	1.705*	.356	< .001	.82	2.59
		Moderate	1.432*	.356	.001	.55	2.31
Item 10	High	Moderate	.182	.315	.834	-.60	.96
		Low	-1.614*	.344	< .001	-2.47	-.76
	Moderate	High	-.182	.315	.834	-.96	.60
		Low	-1.795*	.344	< .001	-2.65	-.94
	Low	High	1.614*	.344	< .001	.76	2.47
		Moderate	1.795*	.344	< .001	.94	2.65
Item 11	High	Moderate	.091	.250	.930	-.53	.71
		Low	-1.364*	.272	< .001	-2.04	-.69
	Moderate	High	-.091	.250	.930	-.71	.53
		Low	-1.455*	.272	< .001	-2.13	-.78
	Low	High	1.364*	.272	< .001	.69	2.04
		Moderate	1.455*	.272	< .001	.78	2.13
Item 12	High	Moderate	-.364	.306	.469	-1.12	.39
		Low	-1.795*	.333	< .001	-2.62	-.97
	Moderate	High	.364	.306	.469	-.39	1.12
		Low	-1.432*	.333	< .001	-2.26	-.61
	Low	High	1.795*	.333	< .001	.97	2.62
		Moderate	1.432*	.333	< .001	.61	2.26
Item 13	High	Moderate	-.636	.303	.109	-1.39	.11
		Low	-2.045*	.330	< .001	-2.86	-1.23
	Moderate	High	.636	.303	.109	-.11	1.39
		Low	-1.409*	.330	< .001	-2.23	-.59
	Low	High	2.045*	.330	< .001	1.23	2.86
		Moderate	1.409*	.330	< .001	.59	2.23

Table 4. Tukey's HSD Results(continued)

Dependent Variable	(I) Proficiency level	(J) Proficiency level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Item 14	High	Moderate	.182	.260	.766	-.46	.83
		Low	-1.557*	.284	< .001	-2.26	-.85
	Moderate	High	-.182	.260	.766	-.83	.46
		Low	-1.739*	.284	< .001	-2.44	-1.04
	Low	High	1.557*	.284	< .001	.85	2.26
		Moderate	1.739*	.284	< .001	1.04	2.44
Item 15	High	Moderate	-.455	.258	.201	-1.09	.18
		Low	-2.477*	.281	< .001	-3.17	-1.78
	Moderate	High	.455	.258	.201	-.18	1.09
		Low	-2.023*	.281	< .001	-2.72	-1.33
	Low	High	2.477*	.281	< .001	1.78	3.17
		Moderate	2.023*	.281	< .001	1.33	2.72

* The mean difference is significant at the 0.05 level.

As shown in Table 5, responses to Items 1–5 regarding the first theme (learner familiarity) indicated that the more years of experience teachers had, the more positive their responses were. The first item in this theme was "I think learners' familiarity with AI-driven tools can positively influence their L2 writing with short paragraphs." Participants with more experience were more likely to agree or strongly agree with this item (M = 1.27), followed by the group with moderate experience (M = 1.82), while the less experienced group was more likely to disagree (M = 4).

Table 5. Responses to Item 1

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.27	.647	.195
Moderate	11	1.82	.751	.226
Low	8	4.00	.756	.267
Total	30	2.20	1.324	.242

As shown in Table 6, regarding Item 2 (I think learners' familiarity with AI-driven tools can positively influence their L2 writing with long paragraphs), the more experienced group responded more positively (M = 1.36), followed by the moderately experienced group, which more often responded with "not sure/neutral" (M = 2.09), while less experienced teachers reported more negative perceptions (M = 3.63).

Table 6. Responses to Item 2

Experience	N	M	SD	Std. Error
High	11	1.36	.505	.152
Moderate	11	2.09	.701	.211
Low	8	3.63	.916	.324
Total	30	2.23	1.135	.207

As shown in Table 7, more experienced (M = 1.36) and moderately experienced (M = 1.82) teachers were more likely to agree with Item 3 (I think learners' familiarity with AI-driven tools can positively influence their L2 writing with argumentative essays) compared to the less experienced group, which was more likely to disagree (M = 3.88).

Table 7. Responses to Item 3

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.36	.505	.152
Moderate	11	1.82	.603	.182
Low	8	3.88	.641	.227
Total	30	2.20	1.186	.217

As shown in Table 8, the moderately experienced group was more likely (M = 1.55) to agree with Item 4 (I think learners' familiarity with AI-driven tools can positively influence their L2 writing with persuasive essays) than the more experienced group (M = 1.73), while the less experienced group was again more likely to disagree (M = 3.63).

Table 8. Responses to Item 4

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.73	.786	.237
Moderate	11	1.55	.522	.157
Low	8	3.63	.744	.263
Total	30	2.17	1.117	.204

As shown in Table 9, the more experienced group again had the highest agreement (M = 1.36) with Item 4 (I think learners' familiarity with AI-driven tools can positively influence their intrinsic motivation toward L2 writing), followed by the moderately experienced (M = 1.64) and less experienced (M = 3.75) groups.

Table 9. Responses to Item 5

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.36	.505	.152
Moderate	11	1.64	.674	.203
Low	8	3.75	1.035	.366
Total	30	2.10	1.242	.227

The second theme, teacher familiarity, had five items, the first of which was "I think instructors' familiarity with AI-driven tools can positively influence teaching L2 writing with short paragraphs." As shown in Table 10, the most positive responses were from the more experienced group (M = 1.27), followed by the moderately (M = 1.36) and less (M = 3.38) experienced groups.

Table 10. Responses to Item 6

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.27	.467	.141
Moderate	11	1.36	.505	.152
Low	8	3.38	.916	.324
Total	30	1.87	1.106	.202

As shown in Table 11, for Item 7 (I think instructors' familiarity with AI-driven tools can positively influence teaching L2 writing with long paragraphs), the highest positive responses were from the more experienced group (M = 1.45), followed by the moderately (M = 1.82) and less (M = 2.88) experienced groups.

Table 11. Responses to Item 7

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.45	.688	.207
Moderate	11	1.82	.405	.122
Low	8	2.88	.835	.295
Total	30	1.97	.850	.155

As shown in Table 12, regarding Item 8 (I think instructors' familiarity with AI-driven tools can positively influence teaching L2 writing with argumentative essays), the highly (M = 1.55) and moderately (M = 1.64) experienced groups had similar levels of agreement, while the less experienced group was more likely to disagree with this statement (M = 3.38).

Table 12. Responses to Item 8

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.55	.522	.157
Moderate	11	1.64	.674	.203
Low	8	3.38	.744	.263
Total	30	2.07	1.015	.185

As shown in Table 13, the more experienced group (M = 1.55) responded more positively than the moderately (M = 1.82) and less (M = 3.25) experienced groups to Item 9 (I think instructors' familiarity with AI-driven tools can positively influence teaching L2 writing with persuasive essays).

Table 13. Responses to Item 9

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.55	.522	.157
Moderate	11	1.82	.751	.226
Low	8	3.25	1.035	.366
Total	30	2.10	1.029	.188

As shown in Table 14, the moderately experienced group was more likely (M = 1.45) to agree with Item 10 (I think instructors' familiarity with AI-driven tools can positively influence teaching motivation toward L2 writing) than the more experienced (M = 1.64) or less experienced (M = 3.23) groups.

Table 14. Responses to Item 10

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.64	.809	.244
Moderate	11	1.45	.522	.157
Low	8	3.25	.886	.313
Total	30	2.00	1.050	.192

As shown in Table 15, the moderately experienced group was slightly more likely (M = 1.55) than the more experienced group (M = 1.64) to agree with Item 11 (I think it's beneficial for L2 high level learners to include AI-driven tools in L2 writing textbooks to improve English writing skills), while the less experienced group was far less likely to agree (M = 3).

Table 15. Responses to Item 11

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.64	.505	.152
Moderate	11	1.55	.522	.157
Low	8	3.00	.756	.267
Total	30	1.97	.850	.155

As shown in Table 16, the more experienced group was more likely (M = 1.45) to agree with Item 12 (I think it's beneficial for L2 moderate level learners to include AI-driven tools in L2 writing textbooks to improve English writing skills) than the moderately (M = 1.82) and less (M = 3.25) experienced groups.

Table 16. Responses to Item 12

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.45	.522	.157
Moderate	11	1.82	.603	.182
Low	8	3.25	1.035	.366
Total	30	2.07	1.015	.185

As shown in Table 17, the more experienced group was more likely (M = 1.45) than the moderately (M = 2.09) and less (M = 3.50) experienced groups to agree with Item 13 (I think it's beneficial for L2 low level learners to include AI-driven tools in L2 writing textbooks to improve English writing skills).

Table 17. Responses to Item 13

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.45	.688	.207
Moderate	11	2.09	.701	.211
Low	8	3.50	.756	.267
Total	30	2.23	1.073	.196

As shown in Table 18, the moderately experienced group was more likely (M = 1.64) than the more (M = 1.82) and less (M = 3.38) experienced groups to agree with Item 14 (I think including AI-driven tools in L2 writing assessment can improve English writing skills).

Table 18. Responses to Item 14

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.82	.751	.226
Moderate	11	1.64	.505	.152
Low	8	3.38	.518	.183
Total	30	2.17	.950	.173

As shown in Table 19, the more experienced group was more likely (M = 1.27) to agree with Item 15 (I think including AI-driven tools in L2 writing assessment can improve English writing pedagogy) than the moderately (M = 1.73) and less (M = 3.75) experienced groups.

Table 19. Responses to Item 15

Experience	*N*	*M*	*SD*	Std. Error
High	11	1.27	.467	.141
Moderate	11	1.73	.647	.195
Low	8	3.75	.707	.250
Total	30	2.10	1.185	.216

3.2 Qualitative Results

The open-ended question (Item 16) asked, "Overall, do you think that AI-driven tools can help improve the writing skills and academic writing of Saudi EFL learners? Specifically, is it possible to provide improvement for learners, pedagogy, textbooks, or assessment practices? Why or why not? Support your point of view, provide examples." The responses revealed some initial confusion, since the interviewer had to explain what "AI-driven tools" meant by providing examples such as ChatGPT and Silatus to Participants 5, 7, 21, and 24. Less than half of participants (46.66%) saw such tools as having a positive effect on education. They also noted the challenges of using these tools in practice, such as a lack of relevant professional training. Participants 1, 3–7, 13, 18–20, 23, and 27–29 strongly agreed that such tools could improve L2 writing. For instance, Participant 1 said they "absolutely support" such tools; Participant 4 said, "I do suggest [using them], but I'm really afraid of improper application"; Participant 20 viewed them as promising and said they would "be fine with some training of both learners and instructors"; and Participant 29 assumed that AI-driven tools would improve EFL writing but with limited improvement and that "advanced learners will get the bigger slice of cake." That participant suggested that learners with lower language performance would depend too much on those tools to do their work for them, leading to lower improvement in their writing skills.

Participants 8–12, 14–16, 22, and 24–26 (40% of the sample) expected that AI integration would have limited-to-no positive impact on the writing skills of Saudi EFL learners. Participant 11 viewed writing as an interactive, collaborative process carried out by learners and teachers. Participants 11, 16, and 25 reported that they could not imagine AI-driven tools improving writing skills instead of being used unethically to generate learners' writing assignments.

Finally, 13.33% of the sample (Participants 2, 17, 21, and 30) did not mention any positive or negative impact of AI integration. They claimed that because AI-driven tools were an external instrument for paraphrasing, summarizing, or generating linguistic output, they had nothing to do with language improvement.

4. Discussion

The first five survey items concerned learner familiarity with AI tools. Regarding Item 1, participants with over 10 years of experience generally agreed that familiarity with AI tools could improve writing skills (see Figure 1). Those with 5–10 years of experience also generally agreed with the item but to a lesser degree. The group with less than five years of experience was more likely to disagree with the statement. This pattern of higher agreement with higher experience was apparent throughout this theme and the rest of the data, with few exeptions. In those exceptions, such as Item 4 (persuasive essays), the group with moderate experience showed slightly higher agreement with the item than the more experienced group, while the less experienced group still tended to disagree. This trend matched the findings of several studies (e.g., Abdelhalim, 2024; Alsahil, 2024; Daweli, 2018; Khan & Kumar, 2023).

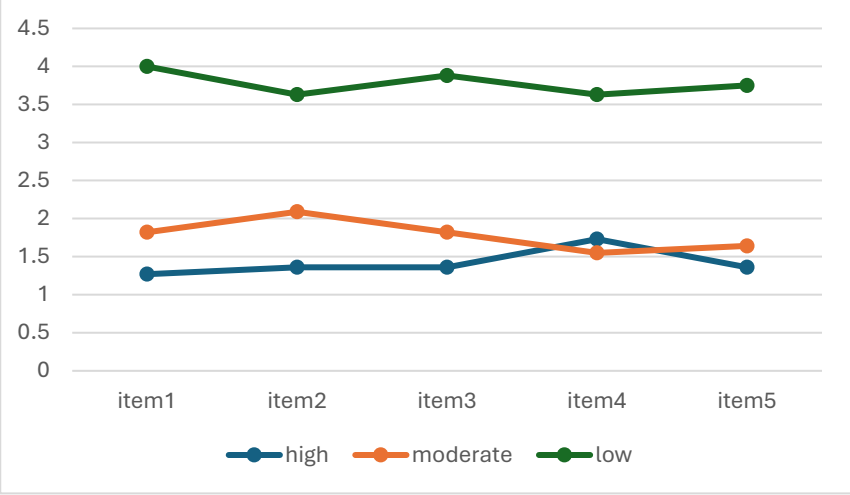

Figure 1. Responses Related to Theme 1

Items 6–10 concerned the second theme, the effect of teacher familiarity with AI tools on teaching how to write short paragraphs, long paragraphs, argumentative essays, and persuasive essays, as well as learner motivation (see Figure 2). Again, the responses showed higher acceptance among participants with over five years of experience teaching. This agreed with previous studies that found a strong correlation between learners' attitudes and practices (Mohammed & Ali, 2021) and that students with an academic major related to computer technology used more language learning strategies (McMullen, 2009).

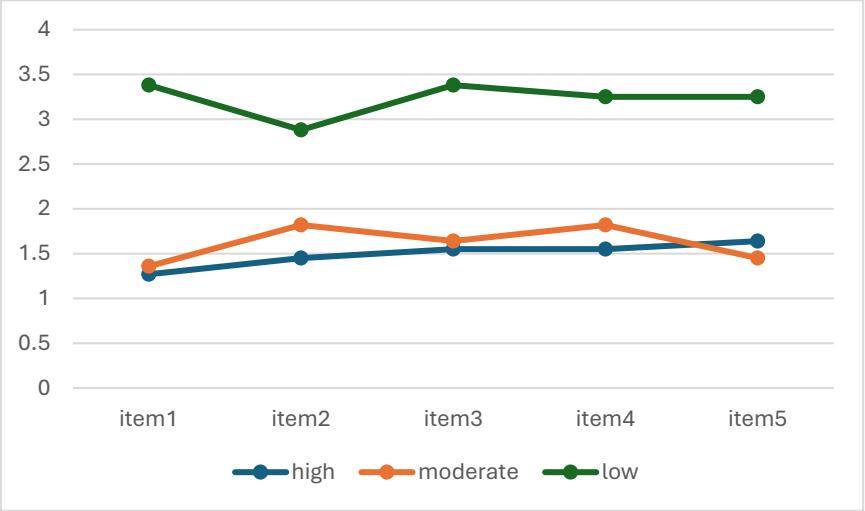

Figure 2. Responses Related to Theme 2

The next three items, related to learning materials, asked about the impact of AI-driven tools on Saudi EFL learners with different levels of writing performance (see Figure 3). More experienced participants had more positive perceptions of such integration regardless of learner performance level. This supported previous findings (e.g., Kwon et al., 2023; Lee et al., 2024).

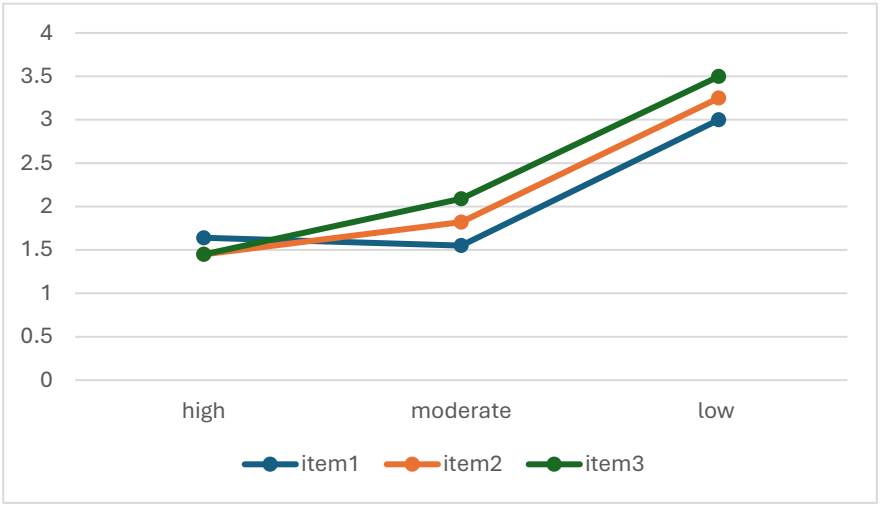

Figure 3. Responses Related to Theme 3

The last theme (Items 14 and 15) was related to the impact of AI and technology tools on EFL assessment (see Figure 4). The first item was "I think including AI-driven tools in L2 writing assessment can improve English writing skills," and the other was "I think including AI-driven tools in L2 writing assessment can improve English writing pedagogy."

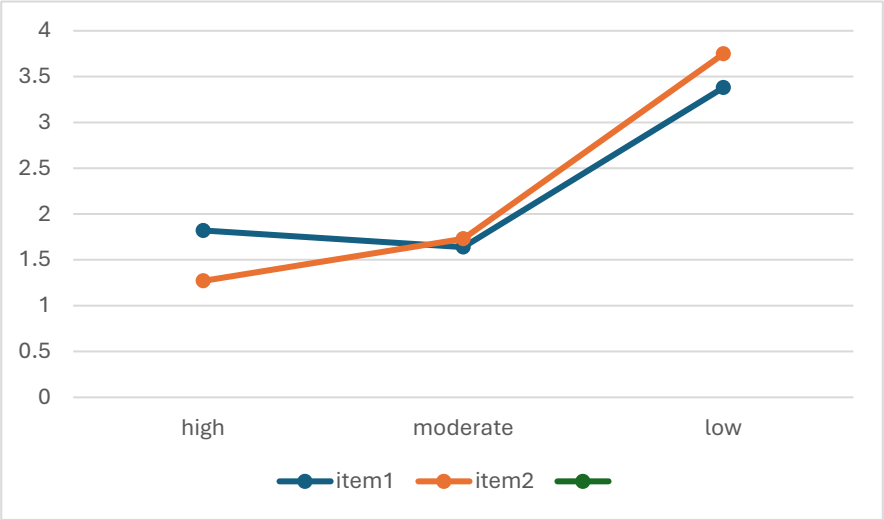

Figure 4. Responses Related to Theme 4

5. Conclusion

In answer to the first research question, the results revealed years of teaching experience as a crucial factor in teachers' acceptance of using AI tools to improve the writing skills of Saudi EFL learners. Teachers with more than 10 years of experience almost always had the most positive views about this technology across the four themes (learner familiarity with AI, teacher familiarity, learning materials, and assessment practices), although those with 5–10 years of experience were often not far behind and sometimes showed even higher agreement with certain items. Teachers with the least experience were far less likely to accept this technology, possibly due to their limited understanding of it, concerns about the tools distracting or confusing learners, the need for advanced training, and unequal access to these tools among instructors and learners.

Regarding the second research question, this study offers several suggestions to improve the writing skills of Saudi EFL learners with AI-driven tools. First, tools could be described in terms of how they can help students rather than focusing on whether they are "AI." Second, students should be guided in their use to avoid confusion or misuse. Third, instructors as well as students would benefit from training before they are asked to use these tools in practice. Fourth, these tools can be used to facilitate writing and support traditional approaches rather than replacing them. Fifth, institutions should ensure that any acceptable AI tools are available to all students equally. Finally, teachers and students need to understand the procedures and objectives for using these tools.

References

Abdelhalim, S. M. (2024). From traditional writing to digital multimodal composing: Promoting high school EFL students' writing self-regulation and self-efficacy. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2024.2322148

Aldosemani, T., Assalahi, H., Lhothali, A., & Albsisi, M. (2023). Automated writing evaluation in EFL contexts: A review of effectiveness, impact, and pedagogical implications. *International Journal of Computer-Assisted Language Learning and Teaching*, *13*(1). https://doi.org/10.4018/ijcallt.329962

Alhaider, S. M. (2023). Teaching and learning the four English skills before and during the COVID-19 era: Perceptions of EFL faculty and students in Saudi higher education. *Asian-Pacific Journal of Second and Foreign Language Education*, *8*(1), 19. https://doi.org/10.1186/s40862-023-00193-6

Alkodimi, K. A., & Al-Ahdal, A. (2021). Strategies of teaching writing at Saudi tertiary-level institutions: Reality and expectations. *Arab World English Journal*, *12*(2), 399-413. https://doi.org/10.24093/awej/vol12no2.27

Almusharraf, A., & Bailey, D. (2023). Machine translation in language acquisition: A study on EFL students' perceptions and practices in Saudi Arabia and South Korea. *Journal of Computer Assisted Learning*, *39*(6), 1988-2003. https://doi.org/10.1111/jcal.12857

Alsahil, A. (2024). Exploring students' perceptions and affordances of Google Docs-supported collaborative writing.

Innovation in Language Learning and Teaching, *19*(1). https://doi.org/10.1080/17501229.2024.2326030

Alshammari, H. A. (2022). Investigating the low English proficiency of Saudi EFL learners. *Arab World English Journal*, *13*(1), 129-144. https://doi.org/10.24093/awej/vol13no1.9

Chen, Y. T., Li, M., Huang, C. Q., Cukurova, M., & Ma, Q. (2024). A systematic review of research on immersive technology-enhanced writing education: The current state and a research agenda. *IEEE Transactions on Learning Technologies*, *17*, 919-938. https://doi.org/10.1109/tlt.2023.3341420

Daweli, T. W. (2018). Engaging Saudi EFL students in online peer review in a Saudi university context. *Arab World English Journal*, *9*(4), 270-280. https://doi.org/10.24093/awej/vol9no4.20

Khan, R. M. I., & Kumar, T. (2023). Metacognitive strategies use in fostering EFL learners' writing skill during remote learning. *International Journal of Innovation and Learning*, *33*(2), 252-268. https://doi.org/10.1504/ijil.2023.128872

Khan, R. M. I., Shahbaz, M., Kumar, T., & Khan, I. (2020). Investigating reading challenges faced by EFL learners at elementary level. *Register Journal*, *13*(2), 277-292. https://doi.org/10.18326/rgt.v13i2.277-292

Kwon, S. K., Shin, D., & Lee, Y. (2023). The application of chatbot as an L2 writing practice tool. *Language Learning & Technology*, *27*(1), 73541. https://doi.org/10.64152/10125/73541

Lee, Y. J., Davis, R. O., & Lee, S. O. (2024). University students' perceptions of artificial intelligence-based tools for English writing courses. *Online Journal of Communication and Media Technologies*, *14*(1), e202412. https://doi.org/10.30935/ojcmt/14195

McMullen, M. G. (2009). Using language learning strategies to improve the writing skills of Saudi EFL students: Will it really work? *System*, *37*(3), 418-433. https://doi.org/10.1016/j.system.2009.05.001

Mohammed, G. M. S., & Ali, J. K. M. (2021). Informal digital learning of English vocabulary: Saudi EFL learners' attitudes and practices. *Arab World English Journal*, *Special Issue on CALL,* (7), 345-358. https://doi.org/10.24093/awej/call7.24

Muftah, M. (2023). Data-driven learning (DDL) activities: Do they truly promote EFL students' writing skills development? *Education and Information Technologies*, *28*(10), 13179-13205. https://doi.org/10.1007/s10639-023-11620-z

Riswanto, Teferi, H., & Ibrahim, K. A. A. (2023). Cultivating EFL learners' productive skills by employing dynamic and non-dynamic assessments: Attitude in focus. *Language Testing in Asia*, *13*(1), 18. https://doi.org/10.1186/s40468-023-00228-2

Sherkuziyeva, N., Gabidullina, F. I., Ibrahim, K., & Bayat, S. (2023). The comparative effect of computerized dynamic assessment and rater mediated assessment on EFL learners' oral proficiency, writing performance, and test anxiety. *Language Testing in Asia*, *13*(1), 15. https://doi.org/10.1186/s40468-023-00227-3

Appendix A

Survey of Instructors' Perceptions of Using AI to Improve the Writing Skills of Saudi EFL Learners

No.	Item	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
1	I think learners' familiarity with AI-driven tools can positively influence their L2 writing with short paragraphs.					
2	I think learners' familiarity with AI-driven tools can positively influence their L2 writing with long paragraphs.					
3	I think learners' familiarity with AI-driven tools can positively influence their L2 writing with argumentative essays.					
4	I think learners' familiarity with AI-driven tools can positively influence their L2 writing with persuasive essays.					
5	I think learners' familiarity with AI-driven tools can positively influence their intrinsic motivation toward L2 writing.					
6	I think instructors' familiarity with AI-driven tools can positively influence teaching L2 writing with short paragraphs.					
7	I think instructors' familiarity with AI-driven tools can positively influence teaching L2 writing with long paragraphs.					
8	I think instructors' familiarity with AI-driven tools can positively influence teaching L2 writing with argumentative essays.					
9	I think instructors' familiarity with AI-driven tools can positively influence teaching L2 writing with persuasive essays.					
10	I think instructors' familiarity with AI-driven tools can positively influence teaching motivation toward L2 writing.					
11	I think it's beneficial for L2 high level learners to include AI-driven tools in L2 writing textbooks to improve English writing skills.					
12	I think it's beneficial for L2 moderate level learners to include AI-driven tools in L2 writing textbooks to improve English writing skills.					
13	I think it's beneficial for L2 low level learners to include AI-driven tools in L2 writing textbooks to improve English writing skills.					
14	I think including AI-driven tools in L2 writing assessment can improve English writing skills.					
15	I think including AI-driven tools in L2 writing assessment can improve English writing pedagogy.					

Open-Ended Question: Overall, do you think that AI-driven tools can help improve the writing skills and academic writing of Saudi EFL learners? Specifically, is it possible to provide improvement for learners, pedagogy, textbooks, or assessment practices? Why or why not? Support your point of view, provide examples.

Acknowledgments

Not applicable.

Authors contributions

Not applicable.

Funding

This work was funded by the Deanship of Graduate Studies and Scientific Research at Jouf University under grant No. (DGSSR-2024-03-02155)

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.