International Journal of TESOL Studies (2026)
260702, 1-27 https://doi.org/10.58304/ijts.260702

Article

Voice and Ownership in AI-Assisted L2 Writing

Aser Altalib
College of Arts, Jouf University, Saudi Arabia

Received: 1 May 2026/Received in revised form: 22 June 2026/Accepted: 24 June 2026/
Available online: 10 July 2026

Abstract
Generative artificial intelligence (AI) has unsettled a long-standing premise of second language (L2) writing research: that the language through which a writer's voice is realised originates with the learner. Voice (what a text conveys about its writer) and psychological ownership (whether writers feel the text is theirs) capture different aspects of a writer's relationship to a text, yet have been theorised separately and rarely examined together. Using a sequential explanatory mixed-methods design, this study operationalised voice and ownership together in one sample of Saudi EFL undergraduates, most in their first year and new to AI, a population rarely examined in identity-focused research on AI-assisted writing. A questionnaire was completed by 178 students; 17 were interviewed using maximum variation sampling. Data were analysed using descriptive and inferential statistics and reflexive thematic analysis. AI use was assistance-oriented rather than generative, supportive use exceeding full-text generation (d = 0.86). Perceived voice was near the midpoint while ownership fell significantly below it; the two were strongly related (r = .608) yet not redundant. Supportive use was the strongest predictor of both perceptions, whereas generative use predicted ownership but not voice, and proficiency made a small, tentative contribution to voice only. Interviews showed ownership to be a graded judgement turning on transformation and accountability, and voice to rest on recognisability and level-match; full-text generation occupied the authorship boundary. These findings suggest pedagogy and assessment should focus less on whether AI is present than on whether learners understand, transform, and authorise its contribution.

Keywords
Generative AI, AI-assisted writing, L2 writing, authorial voice, psychological ownership

1 Introduction

Few questions in applied linguistics have proven as durable as the question of how a writer's identity is realised in the text they produce. From Ivanič's (1998) account of the discoursal self through Hyland's (2005, 2012) interactional model of stance and engagement, and Matsuda's (2001) cross-linguistic work on voice, the field has reached a near-consensus that writing is, among other things, an act of self-

Email: aaltalib@ju.edu.sa

representation. Through that act, the writer projects a recognisable self into the text. In the L2 context, this projection has long been understood to be both pedagogically central and analytically fraught (e.g., Hyland, 2011; Tardy, 2012; Zhao, 2013). The construct has accordingly been operationalised through analytic rubrics, corpus frameworks, and discoursal-identity models that, despite their differences, share one foundational assumption: that the language through which voice is realised originates, however imperfectly, with the learner. This assumption has rarely been questioned because, until recently, it has rarely needed to be.

The widespread availability of generative AI tools to L2 writers has made this assumption much harder to sustain. When a learner submits an essay in which sentences have been generated by a large language model, paraphrased by an AI rewriter, and stylistically upgraded by a grammar tool, the question of whose voice is being measured, and on what basis it is being attributed to the learner, is no longer only a theoretical concern. This is not only a question about the writer but about the construct of voice itself: if voice is the textual trace of identity work, what happens when part of that text is produced by a system with no identity of its own, and what are raters, corpus measures, or self-report items then capturing? The early empirical literature has developed quickly but unevenly, first around the affordances of these tools for specific writing tasks (e.g., Barrot, 2023; Yan, 2023) and then around writing-quality outcomes (e.g., Mizumoto et al., 2024; Shi et al., 2025). Less developed is the strand asking what AI-assisted writing does to the learner's relationship to the text — which, as Sandstead and Kibler (2025) argue, should place voice at the centre of L2 writing research in the AI era.

Sandstead and Kibler's (2025) call is important, but on its own it is incomplete. Voice describes what the text does; it does not necessarily describe what the writer feels about the text. The two can come apart in ways that generative AI makes more visible: a text may carry surface markers of authorial presence that satisfy a rater's rubric while leaving its nominal author feeling that the text is, in some important sense, not theirs. This is where psychological ownership (Pierce et al., 2003) enters as a complementary lens, recently operationalised for AI-assisted writing in particular (e.g., Draxler et al., 2024; Joshi & Vogel, 2025; Kosmyna et al., 2025). This body of work suggests that the felt authorship of an AI-assisted text may be weakened even where its surface qualities are not, raising a further question: can voice and ownership be measured together in the same sample, and if so, what does their relationship reveal about the experience of writing with AI? To the author's knowledge, the two constructs have not yet been operationalised together in an empirical L2 writing study, despite the conceptual case for doing so.

A second gap concerns where the emerging account of AI-era L2 writer identity is taking shape. The identity-focused strand has, to date, been conducted predominantly in East Asian and Western contexts, while comparable work from Saudi Arabia and the wider Gulf has remained largely quantitative and acceptance-focused. It is unclear whether the identity dynamics documented elsewhere extend cleanly to a population whose conditions of AI-assisted English writing differ in important ways from those in the current evidence base. The present study aims to address both gaps by operationalising voice and ownership together, in the same Saudi EFL undergraduate sample, through a sequential explanatory mixed-methods design. Its primary contribution is conceptual: it pairs two constructs the field has so far developed in parallel and asks what their relationship reveals about how learners negotiate the boundary between AI assistance and authorship. Its setting — Saudi EFL undergraduates, a population the identity-focused literature has scarcely examined — provides the empirical site for that question rather than the contribution itself. The answer bears directly on how L2 writing is taught and assessed when AI is in routine use: on what instruction should ask learners to do, and on what assessment can reasonably require them to demonstrate.

2 Literature Review

2.1 Voice, identity, and the writer's self in L2 writing

Voice rests on a premise the field now treats as near-settled: that writing is an act of self-representation rather than the simple transfer of information. Writers do not only report ideas; they also project a sense of self. In the L2 context, this projection becomes especially complex. Ivanič (1998) argued that learners invest a 'discoursal self' into the texts they produce, a self shaped by the resources available to them and by the 'possibilities for self-hood' offered by the target discourse community. Voice, on this view, is not a matter of style alone but the textual trace of the writer's identity work.

Hyland's (2005; see also Hyland, 2011) interactional framework operationalised this insight by showing that voice is realised through patterned choices in hedges, boosters, self-mention, and engagement markers, and that these choices are both personal and disciplinary: writers signal who they are while signalling where they belong. When a writer chooses 'it can be argued that' over 'the data show', the choice is not only stylistic; it positions the writer within a discourse community. Subsequent work has often observed that L2 writers tend to minimise authorial presence relative to L1 peers, producing texts that read as linguistically competent but rhetorically muted (e.g., Hyland, 2011). The pedagogical implication has been clear: developing L2 writers' voice has long been treated as a core aim of writing instruction.

Alongside this theoretical work, a parallel literature has sought to measure voice more directly. Whether 'individualised voice' contributes to L2 writing quality was initially doubted (Helms-Park & Stapleton, 2003) but later supported through analytic rubrics distinguishing authorial presence, stance, and ideational expression as separable dimensions that can correlate with raters' holistic judgements (Yoon, 2017; Zhao, 2013; Zhao & Wu, 2022). Tardy (2012) reframes the debate: voice is not a single variable that is simply present or absent, but a multi-dimensional construction varying across individual, social, and dialogic planes.

Despite their differences, these approaches share one assumption that is often left unstated: that the language through which voice is expressed originates, however imperfectly, with the learner. This raises a question that is now unavoidable. If much of the text a learner submits has been generated, paraphrased, or stylistically upgraded by an AI system, what is the voice being measured, and whose is it? If voice is the textual trace of identity work, and the text is co-produced with a system that has neither identity nor work of its own, the construct may need to be considered not only in the text itself but also in the writer's experience of it.

2.2 Psychological ownership and the felt authorship of text

If voice has been a central construct in applied linguistics for theorising the writer's self-in-text, psychological ownership offers a complementary lens that remains under-used in L2 writing research. Pierce et al. (2003) defined psychological ownership as the state in which individuals feel that a target is 'theirs', and theorised three routes through which the feeling develops: exercising control over the target, gaining intimate knowledge of it, and investing the self in it. Among its recognised consequences is felt responsibility for the target, a dimension foregrounded in the Psychological Ownership Questionnaire (Avey et al., 2009). Developed within organisational behaviour, the construct has since been applied to human–AI co-production, including AI-assisted writing (Draxler et al., 2024; Joshi & Vogel, 2025).

Psychological ownership is useful for the AI-writing problem because it separates the feeling of authorship from the fact of production. A writer can feel strong ownership over a text they did not write word for word if they invested substantially in prompting, selecting, and revising it; conversely, a writer who typed every word may feel little ownership if the text was effectively dictated to them. An L2 learner who drafts with AI and then substantially reshapes its argument, examples, and register may end

up with a text that feels more theirs than one they typed unaided but copied closely from a sample paper. Voice and ownership thus capture different things: what the text does, and what the writer feels about it.

Three recent empirical studies make this pairing more plausible. Draxler et al.'s (2024) 'AI ghostwriter effect' study reported that writers who used AI to produce short texts were reluctant to claim authorship even when they had directed and edited the output, pointing to a gap between instrumental use and felt ownership. Joshi and Vogel (2025) measured psychological ownership of AI-assisted writing through a four-item composite (personal ownership, responsibility, personal connection, and emotional connection) and reported high internal consistency across two experiments (α = .92–.96). In a preprint study, Kosmyna et al. (2025) similarly reported lower self-reported ownership among LLM-assisted essay writers, alongside neural evidence that the authors interpret as reduced cognitive engagement. Taken together, these findings suggest that AI assistance may weaken learners' felt authorship of the texts it helps produce — an implication the L2 writing literature has not yet engaged with empirically.

2.3 Generative AI and the reconfiguration of L2 writing

The release of ChatGPT in late 2022 prompted a rapid shift in L2 writing research. Warschauer et al. (2023) frame AI-generated text for L2 writers through three contradictions involving imitation and identity, unequal benefits for already-proficient users, and the risks learners face whether they use AI or avoid it. This framing remains useful because it directs attention to the tensions AI use creates.

Empirical work has grown rapidly since then, but much of it has focused on tool affordances and writing-quality outcomes, as a recent systematic review of ChatGPT for EFL writing documents (Teng, 2024b). Studies have examined AI support for feedback, revision, and idea generation (e.g., Barrot, 2023; Hyland, 2026; Teng, 2024a; Yan, 2023), as well as effects on writing outcomes in quasi-experimental designs (e.g., Mizumoto et al., 2024; Shi et al., 2025; Teng & Huang, 2026). A recurring finding is that AI can improve surface features of L2 writing while raising concerns about over-reliance and learner development. A related strand has examined the metacognitive dimension of AI-assisted writing—how learners plan, monitor, and evaluate their use of ChatGPT for feedback and revision—and has linked metacognitive awareness to learners' writing experiences and performance (Teng, 2025a, 2025b; Teng & Shen, 2025). A smaller strand has begun to examine cognitive and psychological dimensions of AI-assisted writing, including voice, ownership, and agency (e.g., Godwin-Jones, 2024; Kosmyna et al., 2025).

Within this smaller body of identity-focused work, Sandstead and Kibler (2025) argue that voice should be treated as a central theoretical construct for AI-era L2 writing research, not simply as another variable, but as a way of bringing together questions of learner identity, agency, and authorial integrity. Their position is complemented by corpus and comparative work showing differences between AI-generated and student-written texts, including flatter markers of authorial presence in AI-generated writing (Jiang & Hyland, 2025; Nañola et al., 2025). Read together, these contributions suggest a shift in emphasis: the question is not only whether AI improves L2 writing quality, but also what it does to the learner's relationship to the text.

One limitation in this emerging work, however, is that it often treats L2 writers as a relatively undifferentiated group whose interactions with AI are assumed to mirror those of L1 writers. Warschauer et al. (2023) raised this concern directly: the contradictions they identified operate differently for learners whose access to the target language is itself mediated through AI tools in ways that L1 writers' access is not. Yet much empirical work has yet to take up this challenge, and identity-related questions have until recently been addressed mainly in a smaller qualitative literature.

2.4 From perception to identity: The regional absence

This smaller identity-focused literature, concentrated in East Asian and Western contexts, has begun to move from perceptions and performance towards identity and authorship as the analytical frame. AI-assisted paraphrasing has been found to reshape learners' understanding of paraphrasing and to elicit varying degrees of critical engagement, theorised as differential authorial agency (Xu & Zheng, 2025), and AI use has been shown to involve distinct forms of identity negotiation, from reluctant adoption to strategic realignment (Hu et al., 2025). Critical AI literacy has likewise been modelled as a matter of awareness, positionality, strategies, and evaluation (Wang & Wang, 2025), whose interview data include a learner's comment that AI-assisted text 'does not sound like me' — pointing to the dissociation between voice and ownership the present study aims to measure. Darvin's (2025) synthesis argues that generative AI reshapes the conditions under which L2 learners invest in target-language identities.

In the Saudi and Gulf literature, by contrast, work on AI-assisted L2 writing has remained largely quantitative and acceptance-oriented, reporting positive attitudes towards ChatGPT, Technology Acceptance Model predictors of intention to use, and integrity concerns that do not prevent use (e.g., Alghasab, 2025; Alharbi & Al-Ahdal, 2025; Alsofyani & Barzanji, 2025). This work is valuable, but it leaves a specific absence: little is known about how Saudi EFL learners experience AI-assisted writing as a practice tied to identity and authorship — whether they feel the resulting texts are theirs, whether their voice survives the AI's interventions, and on what grounds they separate their own writing from AI-generated text. Recommendations drawn from acceptance-oriented research may also not map neatly onto the conditions under which Saudi undergraduates write, among them Arabic–English bilingualism, discipline-specific English requirements, and institutional AI policies still being formulated.

2.5 The present study

As argued above, the L2 writing literature has approached the identity question posed by generative AI through two largely separate constructs. Voice, as discussed in work by Ivanič (1998), Hyland (2005), and Sandstead and Kibler (2025), theorises the writer's self-in-text. Psychological ownership, theorised by Pierce et al. (2003) and recently applied to AI-writing contexts (Draxler et al., 2024; Joshi & Vogel, 2025; Kosmyna et al., 2025), captures the writer's felt relationship to the text. The field thus has two complementary constructs, each capturing something the other cannot, yet they have not been operationalised together in a single empirical design, and the available AI-writing evidence has come mainly from East Asian and Western contexts. The question this poses is both theoretical and regional: can the voice–ownership pairing be measured together, and if so, what does it reveal about how Saudi EFL undergraduates, a population still under-represented in the identity-focused AI-writing literature, experience the texts they produce with AI help?

The present study aims to address this gap by investigating how Saudi EFL undergraduates perceive their voice and ownership in AI-assisted English writing, and by examining the conditions under which these perceptions strengthen or weaken. More specifically, the study is structured around four research questions:

- **RQ1.** How frequently and for what purposes do Saudi EFL undergraduates use AI tools in English writing?
- **RQ2.** To what extent do learners perceive AI-assisted writing as reflecting their voice and as being their own?
- **RQ3.** To what extent do supportive AI use, generative AI use, and self-rated English proficiency predict perceived voice and ownership?
- **RQ4.** How do learners themselves account for the boundary between their own writing and AI-generated text?

3 Methods

3.1 Research design

The study adopted a sequential explanatory mixed-methods design (Creswell & Plano Clark, 2018), in which a quantitative phase is followed by a qualitative phase that explains and extends the initial findings. In Phase 1, a questionnaire was administered to Saudi EFL undergraduates; in Phase 2, semi-structured interviews were conducted with students who had opted in to a follow-up at the end of the survey. The design was chosen because perceived voice and ownership require both survey measurement and interview-based explanation: the survey captured broad response patterns, while the interviews explored how learners understood terms such as 'my voice' and 'my writing'.

3.2 Participants

Phase 1 participants (N = 178) were undergraduate students enrolled in compulsory English courses at a Saudi public university. Both male and female students were recruited through convenience sampling, with the questionnaire link distributed via institutional email lists, learning management systems, and instructor networks. Participation was voluntary, and no incentives were offered.

Phase 2 interview participants (n = 17; 10 male, 7 female) were drawn from Phase 1 respondents who had indicated at the end of the survey that they were willing to be contacted for a follow-up interview. A maximum variation sampling strategy (Patton, 2015) was employed to ensure diversity across gender, year of study, self-rated English proficiency, and AI use frequency. The final interview sample of 17 is consistent with comparable recent studies of AI-assisted L2 writing (e.g., Xu & Zheng, 2025). Interviews were conducted in person or via video conferencing according to participant preference, with arrangements made in line with Saudi higher-education conventions for gender-appropriate interview settings.

3.3 Instruments

The questionnaire was developed in English, translated into Arabic, and administered in Arabic to the participants. The decision to administer in Arabic was taken to reduce the cognitive load of language-switching, to avoid respondents defaulting to the English version (which would reintroduce L2 reading-comprehension variance into a survey not intended to measure reading), and to align administration with respondents' dominant language.

The instrument comprised four sections. Section 1 collected demographic and background data: gender, year of study, self-rated English proficiency on the CEFR scale (A1–C2), the AI tools used for English writing (e.g., ChatGPT, Copilot, Gemini, Claude, Grammarly, QuillBot), and the duration of AI use. Section 2 (AI Use in L2 Writing) consisted of eight items measuring the frequency of AI use across different writing purposes on a five-point scale from 1 (Never) to 5 (Always). Items were developed for this study, drawing on the conceptualisation of the 'use' dimension in Xu et al. (2025) and the ChatGPT Usage Scale (Sobaih et al., 2024), as well as on the purpose-of-use categories documented in Barrot (2023), Yan (2023), and Xu and Zheng (2025). Because existing scales measure AI literacy or general perceptions rather than purpose-of-use frequency, a frequency-specific measure was constructed. For analysis, the eight items were grouped into two composites reflecting distinct modes of engagement: a supportive AI use composite (idea generation, grammar correction, vocabulary improvement, paraphrasing, Arabic-to-English translation, and academic/formal style improvement) and a generative AI use composite (complete paragraph generation and complete essay/text generation).

Section 3 (Perceived Ownership) comprised six items (one reverse-coded) mapped to the four-dimension framework of psychological ownership in AI-assisted writing reported by Joshi and Vogel

(2025): personal ownership, responsibility, personal connection, and emotional connection. The two personal ownership items were adapted from the Pierce et al. (2003) tradition as operationalised in the Psychological Ownership Questionnaire (Avey et al., 2009). The responsibility item captured the accountability dimension foregrounded in later psychological-ownership measurement work (Avey et al., 2009). The personal-connection and emotional-connection items were directly adapted from Nicholes's (2017) three-item writing-ownership scale (α = .89–.96 in the original validation), which provided verbatim source wording. A reverse-coded counterpart to the personal-ownership item was included to reduce acquiescence bias. All items used a five-point agreement scale from 1 (Strongly Disagree) to 5 (Strongly Agree).

Section 4 (Voice) comprised six items (two reverse-coded) on the same five-point agreement scale. The instrument is presented as a newly constructed learner self-report scale, responding to Sandstead and Kibler's (2025) call to treat voice as a central construct in AI-era L2 writing research, and reflecting the lack of an established learner self-report voice scale for AI-assisted L2 writing. Existing instruments operationalise voice either as a rater-judgement rubric (e.g., Helms-Park & Stapleton, 2003; Yoon, 2017; Zhao, 2013; Zhao & Wu, 2022) or as a corpus-analytic construct realised through patterned engagement and stance markers (e.g., Hyland, 2005; Jiang & Hyland, 2025). Neither approach captures what the present study aims to measure: the writer's own perception of whether their voice remains present in AI-assisted text. Items were therefore grounded in Ivanič (1998) on the discoursal self, Hyland (2005, 2012) on stance and engagement, Matsuda (2001) on voice in L2 writing, and Tardy (2012) on the multi-dimensional nature of voice, and reformulated as first-person self-report statements that ask the writer to judge their own voice in AI-assisted text.

The instrument was translated into Arabic by a bilingual applied linguist and independently back-translated into English by a second bilingual applied linguist who had not seen the original (Brislin, 1970). Discrepancies between the two English versions were resolved through discussion, with items re-translated until equivalence was achieved. The Arabic-only instrument was piloted with undergraduate EFL learners (n = 25) from the same population but not included in the main sample. The pilot was used to check item clarity, estimate completion time, and identify wording problems. Minor wording revisions were made before main administration.

3.4 Interview protocol

The semi-structured interview protocol was developed to address Research Question 4 and to explore participants' explanations for the quantitative patterns addressed in Research Question 3. Eleven guiding questions, seven of which were designated as core, probed participants' AI use practices, their sense of authorship over AI-assisted texts, the conditions under which they felt AI-assisted writing was or was not their own, the line they drew between AI-assisted and AI-generated writing, and the strategies, if any, they used to preserve their voice. Interviews lasted 20–30 minutes and were conducted in Arabic, English, or a mix of both according to participant preference. All interviews were audio-recorded with consent and transcribed verbatim; pseudonyms (P01–P17) replaced any identifying information at the transcription stage. For analysis, the full transcripts were reviewed, and segments bearing on the study's focal constructs of voice and ownership were coded and analysed as described below; the original Arabic was retained throughout.

3.5 Data analysis

Descriptive, reliability, correlation, and regression analyses were conducted using IBM SPSS Statistics, version 29. After data screening, reverse-coded items were recoded and composite scores for supportive AI use, generative AI use, ownership, and voice were computed as scale means. Internal consistency was assessed via Cronbach's α for each composite. Descriptive statistics addressed Research Questions 1

and 2: frequencies and percentages were reported for AI tools used and purposes of use, and means and standard deviations for the voice and ownership composites and their constituent items. Demographic variables (gender, year of study, and CEFR level) were reported descriptively to characterise the sample. Because the voice scale was newly constructed and the ownership scale newly assembled, the twelve voice and ownership items were also submitted to confirmatory factor analysis in lavaan (Rosseel, 2012), with items treated as ordered-categorical and estimated using the WLSMV estimator; a one-factor authorship model was compared with a correlated two-factor model separating voice and ownership.

Research Question 3 was addressed through two parallel multiple regression models (one predicting voice, one predicting ownership) with supportive AI use, generative AI use, and self-rated English proficiency (CEFR, coded from A1 = 1 to C2 = 6 and entered as an ordered numeric predictor) entered simultaneously as predictors. Because CEFR was a single self-rated ordinal descriptor, its treatment was examined in a categorical robustness check reported below. Modelling supportive and generative AI use as distinct predictors, rather than collapsing them into a single use-frequency index, was central to the study's aim of distinguishing how different modes of AI engagement relate to voice and ownership. Gender and year of study were retained for sample description rather than entered into the focal regression models, because the models were designed to test the three predictors specified in Research Question 3. Regression assumptions (linearity, homoscedasticity, normality of residuals, and absence of multicollinearity) were checked through residual diagnostics and variance inflation factors prior to interpretation. Standardised beta coefficients, R^2, and 95% confidence intervals are reported.

Qualitative data were analysed using reflexive thematic analysis (Braun & Clarke, 2019), following the six phases of familiarisation, initial coding, theme generation, review, definition, and write-up. During theme generation and review, the initial codes were organised into a small set of provisional candidate themes, which were then checked against the coded extracts and the research questions; overlapping candidate themes were consolidated, and less central descriptive codes were retained as contextual background rather than developed into themes. Arabic and mixed-language transcript segments selected for reporting were translated into English by the researcher for analytical consistency, and the translations were checked against the original Arabic or source wording by a bilingual applied linguist, with attention to participants' intended meanings rather than literal equivalence; the original-language transcripts were retained throughout so that excerpts could be verified against the source wording. Coding was primarily inductive but analytically informed by the voice and psychological ownership literatures reviewed above. Themes were generated before the strands were integrated, so the analysis was not constructed to mirror the regression model but examined afterwards for how it might explain, complicate, or qualify the quantitative patterns. Consistent with the sequential explanatory design, the qualitative analysis addressed Research Question 4 and also illuminated the quantitative patterns across Research Questions 1–3 through participants' own accounts. A brief reflexive journal was kept after each interview to record what had surprised the researcher and how the researcher's positioning may have shaped the conversation; these notes informed the analytic process at the theme-generation and review stages.

The study received ethical approval before data collection. Participation was voluntary. Survey responses were collected anonymously, while interview data were treated confidentially and pseudonymised at transcription. Participants were informed that their participation or non-participation would have no academic consequences.

4 Results

The results are reported in two phases. Quantitative findings are presented first, addressing Research Questions 1–3. Qualitative findings are then presented to address Research Question 4 and to help interpret the quantitative patterns.

4.1 Quantitative results

4.1.1 Preliminary analysis and sample characteristics

The final quantitative sample comprised 178 students. As Table 1 shows, gender was almost evenly distributed (92 male, 51.7%; 86 female, 48.3%), and most participants were first-year students (n = 144, 80.9%), with the remainder distributed across Years 2 to 5+. Self-rated proficiency on the CEFR scale ranged across all six levels, the largest groups being at B1 (n = 50, 28.1%) and B2 (n = 45, 25.3%). Most respondents were relatively recent adopters of AI for English writing: 79 (44.4%) reported less than six months of use, and a further 53 (29.8%) reported between six months and one year.

For AI use, two composites were computed: supportive AI use and generative AI use. Internal consistency was acceptable for these two composites and for the ownership and voice composites: supportive AI use (α = .816), generative AI use (α = .831, inter-item correlation = .711), ownership (α = .748), and voice (α = .752).

Table 1
Sample Characteristics (N = 178)

Variable	Category	*n*	%
Gender	Male	92	51.7
	Female	86	48.3
Year of study	Year 1	144	80.9
	Year 2	6	3.4
	Year 3	11	6.2
	Year 4	9	5.1
	Year 5+	8	4.5
CEFR proficiency	A1	28	15.7
	A2	38	21.3
	B1	50	28.1
	B2	45	25.3
	C1	6	3.4
	C2	11	6.2
AI-use duration	Less than 6 months	79	44.4
	6–12 months	53	29.8
	1–2 years	26	14.6
	More than 2 years	20	11.2

4.1.2 AI use: Tools and purposes (RQ1)

Research Question 1 asked how frequently and for what purposes Saudi EFL undergraduates use AI tools in English writing. As Table 2 shows, two tools were selected much more frequently than the others: ChatGPT, selected by 140 of the 178 respondents (78.7%), and Gemini, selected by 111 (62.4%); every other tool was selected by fewer than 7% of the sample.

The purpose-of-use means in Table 2 indicate that AI was used most frequently for supportive, form-focused tasks, such as translation, grammar correction, and paraphrasing, and least frequently for

generative tasks. Means ranged from a high for Arabic-to-English translation (M = 3.26, SD = 1.22) to a low for producing complete essays (M = 2.27, SD = 1.03). The two composites reflected this pattern: supportive AI use (M = 3.10, SD = 0.78) was higher than generative AI use (M = 2.30, SD = 0.95), and a paired-samples t-test confirmed the difference as significant, with a large effect, $t(177) = 11.53$, $p < .001$, $d = 0.86$. One-sample tests against the scale midpoint showed that generative use fell significantly below it, $t(177) = -9.88$, $p < .001$, $d = -0.74$, whereas supportive use did not differ significantly from it, $t(177) = 1.74$, $p = .084$. Overall, AI use in this sample was assistance-oriented rather than generative.

Table 2
AI Tools Used and Descriptive Statistics for AI-Use Purposes (N = 178)

Tool used (multiple response)	***n***	**% of sample**
ChatGPT	140	78.7
Gemini	111	62.4
DeepSeek	11	6.2
Copilot	8	4.5
Grammarly	8	4.5
Claude	7	3.9
QuillBot	6	3.4
Other	5	2.8
No listed tool selected	3	1.7
Purpose of use	*M*	SD
Translate Arabic into English	3.26	1.22
Check/correct grammar	3.23	1.10
Make writing more academic/formal	3.13	1.13
Improve vocabulary	3.08	1.07
Paraphrase/reword	3.05	1.04
Generate ideas	2.85	0.92
Generate complete paragraphs	2.33	1.02
Generate complete essays/texts	2.27	1.03
Supportive AI use (composite)	3.10	0.78
Generative AI use (composite)	2.30	0.95

Note. Tool percentages are based on the full sample (N = 178); respondents could select more than one tool. Purpose items were rated 1 (*Never*) to 5 (*Always*). Composite scores are scale means.

4.1.3 Perceived voice and ownership (RQ2)

Research Question 2 asked to what extent learners perceive AI-assisted writing as reflecting their voice and as being their own. The item-level and composite descriptives are reported in Table 3. Perceived voice was close to the scale midpoint (M = 3.06, SD = 0.61) and did not differ significantly from neutral, $t(177) = 1.30$, $p = .195$. The highest-scoring voice item concerned learners' sense that their own way of expressing ideas came through in AI-assisted writing (M = 3.20, SD = 0.95).

Perceived ownership, by contrast, was below the scale midpoint (M = 2.77, SD = 0.62), and a one-sample test confirmed that it sat significantly below the midpoint, $t(177) = -4.93$, $p < .001$, $d = -0.37$.

The item-level pattern clarifies this contrast: the only ownership item to rise above neutral was the felt-responsibility item (M = 3.24, SD = 0.89), whereas the personal-connection (M = 2.75) and emotional-connection (M = 2.59) items fell below it. This could indicate that students more readily accepted responsibility for AI-assisted texts than felt personally or emotionally connected to them. The contrast suggests that students were more hesitant to claim ownership of AI-assisted writing than to recognise some trace of their own voice in it, although neither construct was clearly above the midpoint.

Table 3

Item and Composite Descriptives for Voice and Ownership (N = 178)

Item	*M*	SD
Perceived voice		
The final text sounds like me	3.01	1.08
It reflects my personal style	3.07	1.01
My own way of expressing ideas comes through	3.20	0.95
It makes my English feel less personal (reversed)	3.09	0.88
I recognise my own voice in it	3.10	0.94
It does not sound like my own voice (reversed)	2.89	0.91
Voice total	3.06	0.61
Perceived ownership		
The writing feels like mine	2.68	0.90
I have a high sense of personal ownership	2.58	0.92
I feel responsible for the text	3.24	0.89
The text does not feel like mine (reversed)	2.78	0.97
I feel a personal connection to the text	2.75	0.94
I feel an emotional connection to the text	2.59	1.01
Ownership total	2.77	0.62

Note. Items were rated 1 (*Strongly Disagree*) to 5 (*Strongly Agree*); reversed items are reported after recoding. Totals are scale means. Item wordings are abbreviated.

4.1.4 Measurement structure of voice and ownership

The descriptive contrast above raises a prior measurement question. Because the voice scale was newly constructed and the ownership scale newly assembled from three sources, the claim that the two represent related but distinguishable constructs requires evidence about their factorial structure, not their correlation and differing item means alone. As described in the Method, the twelve items were submitted to confirmatory factor analysis, comparing a one-factor authorship model with a correlated two-factor model separating voice and ownership. Robust fit indices are reported throughout, and the nested models were compared using a scaled difference test (Table 4).

The two-factor model fitted the data significantly better than the one-factor model, $\Delta\chi^2(1) = 59.41$, $p < .001$, and this advantage was visible across the reported indices: the comparative fit index rose from .664 to .741, the Tucker-Lewis index from .589 to .678, the root mean square error of approximation fell from .173 to .153, and the standardised root mean square residual fell from .111 to .103. The improvement indicates that a single authorship dimension does not adequately represent the items. The latent correlation between the two factors was high but well short of unity ($\varphi = .748$, $p < .001$), consistent

with the strong manifest correlation reported above (r = .608) and indicating that the two constructs, while related, are not redundant.

The superiority of the two-factor model should not, however, be read as endorsement of the full instrument. By absolute standards it fitted poorly: the comparative fit index and the Tucker-Lewis index fell well below conventional thresholds, and the root mean square error of approximation remained high. The standardised loadings (Table 5) helped locate the source. Most directly worded indicators loaded substantially on their intended factors (four voice items between .60 and .84; four ownership items between .73 and .76). The weakest was the reverse-worded voice item ("It makes my English feel less personal"), which did not load meaningfully (λ = .06, p = .395); the other reverse-worded indicators loaded modestly (λ = .31 and .42), and the responsibility item showed a lower but significant loading (λ = .45). This pattern points to reverse-wording method strain rather than a substantive failure of either construct, since the misfit concentrated in negatively worded indicators rather than in one content domain.

Diagnostic models supported this interpretation. Allowing the three reverse-worded items' residuals to correlate improved fit ($\Delta\chi^2(3) = 147.56$, $p < .001$), yet the reverse-worded voice item still failed to load ($\lambda = -0.01$), indicating it did not function as a voice indicator at all. A reduced model retaining only the directly worded items (four voice, five ownership) fitted better, although the root mean square error of approximation remained elevated, with a robust comparative fit index of .902, a Tucker-Lewis index of .865, a root mean square error of approximation of .120, and a standardised root mean square residual of .056, and again the two-factor structure outperformed the one-factor alternative ($\Delta\chi^2(1) = 72.06$, $p <$ .001). The distinction between voice and ownership therefore sharpens once the reverse-wording artefact is set aside, though this directly worded model is reported as a diagnostic, not as a replacement for the full instrument, which was retained for the primary analyses.

A final question concerned the internal structure of the ownership composite, given that its felt-responsibility item behaved differently from the others. Across both the full and directly worded models, the responsibility item loaded significantly on the ownership factor (λ = .45 and .47, both $p < .001$), indicating it belongs with the other ownership items even though learners endorsed it more readily; in the full model, the weakest ownership loading was instead the reverse-worded item. A separate exploratory analysis of the ownership items was treated as a follow-up option, to be used only if the confirmatory results showed responsibility-specific strain; because the weakness lay in the reverse-worded item rather than in the responsibility dimension, that analysis was not undertaken. The six-item composite was therefore retained, the responsibility item's divergence reflecting endorsement rather than dimensionality.

Taken together, these analyses provide preliminary structural support for treating voice and ownership as related but distinguishable: the two-factor model was superior on every index and in both the full and reduced item sets, and the ownership composite held together as a single dimension. At the same time, the weak absolute fit of the full instrument, traced to its reverse-worded items, indicates that the scales are best regarded as an initial operationalisation in need of refinement rather than a validated measure, a limitation considered below.

Table 4

Fit Indices for One-Factor and Two-Factor Models of Voice and Ownership (N = 178)

Model	Scaled χ^2	*df*	CFI	TLI	RMSEA	SRMR
One-factor (authorship)	430.87	54	.664	.589	.173	.111
Two-factor (voice, ownership)	356.08	53	.741	.678	.153	.103

Note. N = 178. Items were treated as ordered-categorical and estimated using WLSMV. Scaled χ^2 values are shown; CFI, TLI, and RMSEA are the robust variants. All χ^2 values were significant at $p < .001$. The two-factor model fitted significantly better than the one-factor model, $\Delta\chi^2(1) = 59.41$, $p < .001$. The latent factor correlation in the two-factor model was φ = .748.

Table 5

Standardised Factor Loadings for the Two-Factor Model (N = 178)

Factor	Item	λ
Voice	The final text sounds like me	.72
Voice	It reflects my personal style	.84
Voice	My own way of expressing ideas comes through	.75
Voice	It makes my English feel less personal (reversed)	.06
Voice	I recognise my own voice in it	.60
Voice	It does not sound like my own voice (reversed)	.31
Ownership	The writing feels like mine	.73
Ownership	I have a high sense of personal ownership	.73
Ownership	I feel responsible for the text	.45
Ownership	The text does not feel like mine (reversed)	.42
Ownership	I feel a personal connection to the text	.76
Ownership	I feel an emotional connection to the text	.74

Note. $N = 178$. λ = standardised factor loading. All loadings were significant at $p < .001$ except the reverse-worded voice item "It makes my English feel less personal" (λ = .06, $p = .395$). Reverse-worded items are reported after recoding. Item wordings are abbreviated and correspond to those in Table 3.

4.1.5 Predicting voice and ownership (RQ3)

Research Question 3 asked to what extent supportive AI use, generative AI use, and self-rated proficiency predict perceived voice and ownership. The bivariate correlations among the study variables are reported in Table 6. Supportive AI use was associated with both voice ($r = .324$, $p < .001$) and ownership ($r = .445$, $p < .001$). Generative AI use was not associated with voice ($r = .083$, $p = .269$) but was associated with ownership ($r = .308$, $p < .001$). Self-rated proficiency was unrelated to either construct at the bivariate level (voice $r = .104$, $p = .168$; ownership $r = -.005$). Voice and ownership were themselves strongly, though not perfectly, associated ($r = .608$, $p < .001$); the two thus shared substantial variance while also leaving considerable variance unshared, a pattern the regression models below clarify.

Table 6

Correlations Among Study Variables (N = 178)

Variable	1	2	3	4	5
1. Supportive AI use	—				
2. Generative AI use	.435***	—			
3. Self-rated proficiency	−.164*	−.226**	—		
4. Ownership	.445***	.308***	−.005	—	
5. Voice	.324***	.083	.104	.608***	—

Note. $N = 178$. Following conventional benchmarks, r values around .10, .30, and .50 were interpreted as small, medium, and large, respectively. Coefficients without an asterisk were non-significant (generative–voice $p = .269$; proficiency–ownership $p = .946$; proficiency–voice $p = .168$). * $p < .05$. ** $p < .01$. *** $p < .001$.

Two parallel multiple regression models were then estimated, one for each outcome, with the results reported in Table 7. The model for voice was significant, $F(3, 174) = 8.78$, $p < .001$, $R^2 = .131$. Supportive AI use was the strongest predictor ($\beta = .367$, $p < .001$), whereas generative AI use did not predict voice ($\beta = -.041$, $p = .604$). Although self-rated proficiency was not significantly correlated with voice at the bivariate level ($r = .104$, $p = .168$), it emerged as a small but significant positive predictor in the regression model ($\beta = .155$, $p = .035$), a pattern suggestive of a small suppression-like effect. The model for ownership was also significant, $F(3, 174) = 16.61$, $p < .001$, and explained 22.3% of the variance ($R^2 = .223$). In the ownership model, supportive AI use was again the strongest predictor ($\beta = .391$, $p < .001$); generative AI use, comprising full-text generation only (complete paragraphs and complete essays), made a smaller but significant contribution ($\beta = .159$, $p = .036$); and proficiency was non-significant ($\beta = .095$, $p = .169$). Collinearity diagnostics were satisfactory in both models (VIF 1.06 to 1.27), and no case exerted undue influence (maximum Cook's distance ≤ .085).

Because self-rated proficiency was a single ordinal CEFR descriptor (scored A1 = 1 to C2 = 6) yet entered into the regression as an interval predictor, a robustness check re-estimated the voice model with proficiency recoded into three categories, low (A1–A2), intermediate (B1–B2), and high (C1–C2), with the intermediate group as the reference. The categorical model was also significant, $F(4, 173) = 6.88$, $p < .001$, $R^2 = .137$. The substantive pattern was unchanged: supportive AI use remained the strongest predictor ($\beta = .357$, $p < .001$) and generative AI use remained non-significant ($\beta = -.033$, $p = .680$). The proficiency effect was carried mainly by the lower end of the range, with low-proficiency learners scoring significantly below the intermediate reference group ($\beta = -0.152$, $p = .042$) while the high-proficiency group did not differ significantly from it ($\beta = .052$, $p = .479$). The categorical specification therefore reproduced the small proficiency association observed in the main model and located it primarily in the contrast between lower- and intermediate-proficiency learners, suggesting that the interval treatment of the descriptor did not distort the substantive findings. A further sensitivity check re-estimated both models using directly worded composites only, excluding the three reverse-worded indicators (VOI4, VOI6, and OWN4). The substantive pattern was again unchanged: supportive use remained a significant positive predictor of both outcomes, and full-text generation use remained non-significant for voice but significant for ownership, indicating that the primary findings were not driven by the weaker reverse-worded indicators.

Table 7

Multiple Regression Models Predicting Voice and Ownership (N = 178)

Predictor	*B*	SE B	β	95% CI	*p*
Voice model					
Supportive AI use	0.284	0.061	.367	[0.164, 0.404]	< .001
Generative AI use	−0.026	0.051	−.041	[−0.127, 0.074]	.604
Self-rated proficiency	0.070	0.033	.155	[0.005, 0.135]	.035
Ownership model					
Supportive AI use	0.311	0.059	.391	[0.194, 0.428]	< .001
Generative AI use	0.105	0.050	.159	[0.007, 0.202]	.036
Self-rated proficiency	0.044	0.032	.095	[−0.019, 0.107]	.169

Note. Voice model: $R^2 = .131$, adjusted $R^2 = .116$, $F(3, 174) = 8.78$, $p < .001$. Ownership model: $R^2 = .223$, adjusted $R^2 = .209$, $F(3, 174) = 16.61$, $p < .001$. CI = confidence interval for *B*; β = standardised coefficient.

Taken together, these patterns indicate that voice and ownership were related but not redundant. The two shared substantial variance, yet their predictors diverged: generative AI use predicted ownership but not voice, and self-rated proficiency predicted voice but not ownership. The constructs therefore moved together without being interchangeable.

4.2 Qualitative results

The qualitative phase comprised 17 semi-structured interviews and primarily addressed Research Question 4, though participants' accounts also illuminated the ownership, voice, and predictor patterns. All 17 interviews informed the analysis; the excerpts below were selected to illustrate recurrent patterns and analytically important contrasts, and not every interviewee is quoted. Excerpts appear in English translation, and counts indicate recurrence within the sample, not a formal content-analytic measure.

Routine use clustered in supportive practices: translating from Arabic, correcting grammar, paraphrasing, supplying more academic vocabulary, and simplifying difficult material; several treated the tool as a near-default reference, one calling it 'my encyclopedia' (P01). Full-text generation appeared less often as an everyday practice and more often as the point at which authorship became uncertain, since the same students who used AI comfortably for correction grew hesitant about texts it had written in full.

4.2.1 Mine to the degree I remake it: ownership and the assistance/authorship boundary

Across the interviews, ownership appeared as a graded judgement rather than an all-or-nothing property. What mattered was transformation: how far the writer had originated the idea, understood the material, revised it, and reshaped its language. The more of this work participants described, the more readily they treated the text as theirs. Conversely, copy-paste or whole-text generation shifted authorship back towards the tool.

The most common mechanism was transformation, with editing, rewriting, restyling, or adding one's own ideas treated as the process that turned AI output into something partly one's own. One participant framed the tool as an instrument rather than an author:

> Excerpt 1 (P02, male). Yes, I feel ownership, because the original idea, the research and the organisational effort come from me. The AI only 'polished' the language and improved the phrasing, exactly as a dictionary or a spellchecker does.

The opposite position was equally clear: output taken without intervention belonged to the machine.

> Excerpt 2 (P03, female). When it gives me a text without my changing even a single letter, I see it at that point as a text entirely from the AI.

Between these positions was the most common response: partial, conditional ownership.

> Excerpt 3 (P04, male). I somewhat feel it is mine, but not 100%, because at first it comes from the AI; but after I edit and add to it, it becomes closer to my style and my understanding.

This conditional ownership was grounded in accountability rather than attachment. Participants described making sure they understood every word, verifying information, and rejecting output they could not stand behind, yet seldom described the text as something that expressed them. This asymmetry — responsibility present, emotional connection largely absent — is the qualitative counterpart of the survey pattern.

This also helps explain why a weak sense of ownership sometimes survived under full generation. Because ownership rested on the *idea* and the *request* as much as on the typing, a text the AI wrote in full could still be claimed when its originating idea was the student's: one felt a generated speech was

his because 'I described everything to it and it just wrote' (P17), and another claimed a wholly AI-built presentation without hesitation (P15). Ownership of the idea could therefore outlast the outsourcing of the language, which may be why generative use kept a weak link to ownership while doing nothing for voice.

A smaller group of participants refused the ownership frame altogether, not from insufficient effort but from a rejection of the premise. One participant explained this by treating knowledge as common property:

> Excerpt 4 (P06, male). No, it is not mine, because information will never become anyone's property … if it belonged to anyone, the world would be full of conceited people.

Others located non-ownership in the non-uniqueness of AI output (P07) or treated the claim to ownership as self-importance (P01). These cases mark the limits of the transformation logic: ownership was often linked to effort, but not simply a function of it.

4.2.2 Sounding like me: Voice as recognisability and level-match

If the first theme turns on whether a text is owned, the second turns on whether it sounds like the writer. Participants understood voice less as an abstract authorial identity than as recognisability, the sense that a text stayed within their own, usually simple and familiar, way of putting things. Voice survived when students brought AI output back into that range, and weakened when the language became too advanced, formal, or unlike their own.

Participants described this threat in three main ways. The first was language that felt too formal or academic to be their own:

> Excerpt 5 (P08, female). [It is not mine] when it contains heavily academic and formal vocabulary compared with what I usually use in my writing … this does not reflect my personality or any of my ideas.

The second was a kind of robotic neatness that made AI text recognisable as AI-generated, 'very organised and exposed' in one participant's words (P15). The third was neutrality, the sense that the tool 'has no clear opinion' and 'no fixed values' and so carried no stance to begin with (P10).

In response, participants tried to preserve voice through restyling, simplification, and selective acceptance of AI suggestions. The clearest articulation framed voice as a matter of linguistic choice:

> Excerpt 6 (P13, female). When I reject a sentence the AI suggested and accept another because it resembles my way of speaking, here I am shaping my identity in the text. Helping with the 'template' does not cancel the 'content' that represents me.

For these L2 writers, recognisability was bound up with proficiency in a way that complicates the construct. To say a text 'did not sound like me' was often to say it sat above one's English level, not that it misrepresented one's stance; conversely, 'my style' was glossed as 'simple', 'clear', and 'easy'. Voice here was not only stance projected into discourse but the recognisable trace of a writer's linguistic range. This helps account for the small, regression-only contribution that self-rated proficiency made to voice but not to ownership: a more proficient writer can pull AI output back into a recognisable form, whereas ownership rested on contribution and effort. The diglossic dimension surfaced where one participant rewrote AI output into colloquial Arabic to make it his own (P11).

This pattern is one reason the survey's voice scores clustered around the midpoint: many participants felt their voice could survive AI use, but only on the condition of the reshaping work that not all of them consistently performed.

4.2.3 A double-edged scaffold: Improvement, dependence, and agency over time

The first two themes concern a particular text; the third concerns the writer's developing relationship with the tool. Here participants were divided, often within a single answer: the same tool was described both as something that could improve writing and as something that could replace the writer's own effort.

On one side, it was described as a teacher that 'noticeably develops your skills' and whose explanations were immediately understandable, so it 'does not make you depend on it' (P01). On the other side, it was described as a source of dependence:

> Excerpt 7 (P09, male). It makes me rely on it more, because it writes everything for me, so I am comfortable and do not need to write and tire myself; and that, unfortunately, is wrong.

Participants often resolved the tension by making the outcome contingent on their own stance, one putting it as 'I use it for help, not for everything' (P05); for a few, the pull towards reliance was driven by anxiety about errors rather than convenience (P12). This theme contextualises the first two: students embraced AI as developmental support on the condition that they, not the tool, remained the agent of the writing.

5 Discussion

Taken together, the two strands suggest that students did not experience their AI-assisted writing as a simple choice between human and machine authorship. The qualitative accounts help explain this pattern. Students described translation, correction, paraphrasing, and restyling as practices through which they could remake AI output into something closer to their own ideas, level, and style. Full-text generation, by contrast, occupied a more ambiguous position. It appeared less able to preserve voice when the language no longer sounded recognisably like the learner, even where a thread of ownership survived because the idea or prompt was the student's. This helps to explain why voice and ownership were strongly related but not identical: voice was anchored in recognisability and linguistic fit, whereas ownership rested on effort, idea contribution, and the responsibility of having understood and checked the text. That the two rested on different kinds of work in learners' own accounts — recognisability and level-fit on one side, contribution and accountability on the other — gives the correlation between them its meaning: the strands converge on two related but separable constructs.

5.1 Patterns of AI use (RQ1)

Use was assistance-oriented rather than generative, and supportive use sat only at a moderate level. One possible explanation is that form-focused support, such as translation, grammar correction, and register upgrading, maps onto the immediate demands of writing in English while drawing on Arabic as a resource. This echoes recent Saudi and Gulf work on EFL learners' academic uses of AI (e.g., Alghasab, 2025; Alharbi & Al-Ahdal, 2025; Sobaih et al., 2024), but the present study adds a purpose-differentiated profile in a setting under-represented in identity-focused research. The pattern should also be read in light of the fact that most respondents were relatively recent AI users.

5.2 Recognising voice, withholding ownership (RQ2)

Here the two constructs diverged. Perceived voice sat near the midpoint, whereas perceived ownership fell significantly below it ($d = -0.37$). Learners appeared to recognise a trace of their voice in the text while stopping short of calling it theirs. This was a difference in strength of endorsement, not a full dissociation between constructs; voice and ownership remained strongly related ($r = .608$). The item

pattern sharpens the point: the only ownership item to rise above neutral concerned felt responsibility, while personal and emotional connection fell below it, suggesting that students accepted accountability for AI-assisted texts more readily than they felt attached to them. Such attenuated ownership echoes work in which AI-assisted text weakens psychological ownership even where users still present themselves as authors (Draxler et al., 2024; Joshi & Vogel, 2025; Kosmyna et al., 2025). Conceptually, the result supports treating voice and ownership as related but distinguishable: voice concerns recognisable authorial presence in discourse, whereas psychological ownership concerns the felt relation between a person and a target (Pierce et al., 2003). The measurement analysis is consistent with this reading, in that a two-factor structure fitted better than a single-factor alternative, although the high latent correlation between the factors and the weak absolute fit of the full instrument indicate that the constructs are closely related rather than sharply separated. Put simply, a writer may hear a voice they do not yet feel is theirs.

5.3 What predicts voice and ownership (RQ3)

Research Question 3 asked whether supportive AI use, generative AI use, and self-rated proficiency predict these perceptions. Supportive use was the strongest predictor of both outcomes (voice $\beta = .367$; ownership $\beta = .391$), whereas generative use, here specifically the generation of complete paragraphs and essays, predicted ownership but not voice. The supportive effect is readily explained as transformation work: in correcting, translating, paraphrasing, and restyling, learners convert AI output into something nearer their own ideas, level, and register, and it is this involvement that appears to underwrite both voice and ownership. The asymmetry is harder to interpret. That generative use retained a weak link to ownership while showing no association with voice accords with the idea that ownership can rest on the originating idea and the prompt even when the language is outsourced (cf. Joshi & Vogel, 2025), whereas voice appears less likely to transfer through generation alone and may need to be actively reasserted through revision. Such reassertion preserved one skilled writer's voice in a recent case study (Jacob et al., 2025). These relationships are associational, however, and the cross-sectional design does not establish their direction.

Proficiency produced the study's most counterintuitive result. It was unrelated to voice at the bivariate level ($r = .104$, $p = .168$), yet once AI use was controlled it emerged as a small but significant positive predictor ($\beta = .155$). It did not predict ownership in either analysis. The association with voice became clearer once AI-use variables were controlled, which could mean that more proficient writers are better placed to pull AI output back into a register they recognise as their own. The interviews support that reading, where voice was glossed as level-match and a 'simple', recognisable style. Ownership, resting on contribution and effort rather than on linguistic range, may not track proficiency in the same way. Such a reading also sits with a literature in which the voice–proficiency link has remained unsettled rather than straightforward (e.g., Helms-Park & Stapleton, 2003; Yoon, 2017; Zhao, 2013; Zhao & Wu, 2022). The result should nonetheless be treated with care. It is a small effect resting on a single self-rated item, and is best read as consistent with that unsettled picture rather than as a firm effect.

5.4 Drawing the boundary (RQ4)

The interviews suggest a boundary drawn not by the amount of AI involvement alone, but by two kinds of work. Ownership was a graded judgement turning on transformation and accountability: the more a learner had originated the idea, understood the material, and reshaped the output, the more readily the text was treated as theirs, with copy-paste or unmodified whole-text generation marking the point at which authorship tended to shift back towards the tool (Theme 1). Voice rested on a different criterion. It depended on recognisability and level-match. A text 'did not sound like me' most often when its register was beyond the writer's own level rather than when it misrepresented a stance (Theme 2). These two forms of work also help explain the survey pattern of attenuated ownership alongside a recognisable

voice. The account converges with descriptions of AI-assisted writing as negotiated authorship rather than a human-versus-machine binary (Hu et al., 2025; Xu & Zheng, 2025). One feature was more context-specific. Voice was repeatedly tied to fit with a writer's English level, and one learner's rewriting of AI output into colloquial Arabic points to a diglossic dimension that has received little attention in the predominantly East Asian and Western evidence base. A minority resisted ownership on principle rather than for want of effort, treating information as common property, which marks the limits of any purely effort-based reading.

The same accounts framed the tool as a double-edged scaffold. The resource that develops a writer was also the one that threatens to write in their place, and participants often resolved the tension by making the outcome contingent on staying the agent of the writing (Theme 3). This aligns with accounts of AI as redistributing, rather than removing, learner agency (Darvin, 2025; Godwin-Jones, 2024; Yan, 2023). In this sense, the qualitative boundary mirrors the quantitative pattern, with full-text generation on the authorship boundary. The study therefore extends current work on AI-assisted L2 writing by showing how a population the identity-focused literature has scarcely examined draws the line between assistance and authorship in terms of transformation, accountability, and recognisability.

6 Implications

The clearest implication of these findings is instructional. If voice and ownership are sustained not by the presence or absence of AI but by what learners do with its output, then the pedagogical question is less whether students may use AI than how they are taught to work on what it produces. Supportive use was both the more common mode and the stronger correlate of voice and ownership, and the qualitative accounts locate the reason in the work itself: revising, checking, restyling, and personalising the output. The two constructs, moreover, rest on different things: voice on whether the language still reads as the learner's, ownership on whether the learner contributed to, understood, and can answer for the text. Instruction that secures one will not automatically secure the other. Teaching learners to bring AI output back to their own level and register supports voice, whereas asking them to originate, verify, and take responsibility for the content supports ownership. A practical corollary is that a learner's first language may be treated as a resource for transformation rather than as interference, particularly in multilingual and diglossic settings, though this requires direct investigation.

A second implication concerns assessment and academic integrity. The findings complicate the assumption, common in many current policy discussions, that the appearance of AI in a text is itself the problem. If ownership rests on contribution, understanding, and accountability rather than on whether a tool was involved, then integrity frameworks built mainly around detecting AI presence risk targeting the wrong object. A more defensible approach would make the learner's contribution visible and assessable by asking students to retain and submit their prompts, the suggestions they accepted or rejected, the changes they made, and a short reflective note explaining how the final text represents their own thinking. Such process evidence shifts assessment from whether AI was used to what the learner understood, decided, and can account for — on the present evidence, where ownership is located. Framing the matter this way also avoids treating ordinary supportive use, the dominant mode in this sample, as though it were misconduct. More broadly, the value of AI assistance lies less in the polish of its output than in whether the learner remains its author.

7 Limitations and Future Research

Several limitations should be noted. The quantitative strand was cross-sectional, so the regression results are associational and cannot establish direction. Supportive use may strengthen voice and ownership, but learners who already feel a stronger sense of voice and ownership may also engage AI differently, and

the present design cannot separate the two accounts. Longitudinal and experimental work is needed to test whether transformation-oriented use produces the perceptions reported here or merely accompanies them.

The study also relies on self-report. Frequency of AI use was reported by participants rather than observed, and English proficiency was self-rated on a single CEFR descriptor rather than independently tested. Both measures are subject to the familiar gap between reported and actual behaviour, and the proficiency result in particular should be read with that caution in mind. This self-report limitation may be especially relevant for full-text generation use, which learners may under-report owing to academic-integrity concerns; the resulting restriction of variance could have contributed to its weak association with voice. The full-text generation use measure was also brief, comprising only two items capturing the generation of complete paragraphs and complete essays; while useful as a focused indicator, its narrowness limits what can be inferred about full-text generation use more broadly. Independent proficiency measures and behavioural records of AI use would strengthen future designs.

The sample was also context-bound. Participants were Saudi EFL undergraduates, most in their first year (80.9%) and many relatively new to writing with AI, so their assistance-oriented profile may reflect that early stage rather than a more general pattern. How voice and ownership behave among more advanced, postgraduate, professional, or L1 writers remains open, and replication would show how far the present account travels beyond its setting. The findings also offer a snapshot of a fast-moving target: the available tools and the purposes learners put them to are shifting quickly, so the usage profile in particular is best read as time-bound.

A related limitation concerns the newly constructed voice measure. Because no established learner self-report scale for voice in AI-assisted writing was available, it was constructed for the present study. Its internal consistency was acceptable, and a two-factor measurement model separating voice from ownership received preliminary support; however, the full twelve-item instrument fitted the data poorly in absolute terms, with weakness concentrated in the reverse-worded items, one of which did not function as intended. Broader convergent and discriminant validity evidence, as well as evidence of performance across other populations, therefore remains to be established. The findings should be read as an initial operationalisation of perceived voice rather than as findings based on a fully validated scale, and the instrument would benefit from refinement, particularly of its reverse-worded items, before reuse.

Finally, the study measured perceived voice and ownership, not the textual features through which voice is conventionally analysed. The study can show what learners believe about their AI-assisted writing, but not whether a reader or a rubric would find the same voice in the text, and that gap between perceived and realised voice is itself worth investigating. The most informative future designs would combine self-report and interviews with the writing itself — drafts, prompt histories, revision traces, and rater or corpus analysis of the final texts — so that perception and product can be compared directly. A further line concerns learners' multilingual repertoires in transforming AI output, which merit dedicated attention in diglossic and multilingual settings.

8 Conclusion

This study set out to examine how Saudi EFL undergraduates, most in their first year and new to AI, use AI in their English writing and how far they regard the resulting texts as carrying their voice and belonging to them. Its contribution is to bring voice and psychological ownership, two constructs rarely studied together, into a single account of AI-assisted L2 writing, in a context that work on AI and writer identity has scarcely examined, and to offer an initial self-report operationalisation of learner-perceived voice that subsequent work can refine and validate. The central claim is that voice and ownership are related but separable: they are supported by the same transformation work, yet anchored differently, voice in recognisability and level-fit, ownership in contribution, accountability, and understanding.

The pattern beneath the findings is consistent. Learners used AI far more to support their writing than to generate it wholesale, and this supportive, transformative use was most closely tied to both voice and ownership. They recognised a trace of their own voice in AI-assisted text more readily than they were willing to call it theirs. Full-text generation sat at the boundary: voice was difficult to preserve once the language ceased to sound like them, even where a thread of ownership persisted because the idea was the learner's.

The central issue in AI-assisted L2 writing is therefore not simply whether AI appears in a text, but whether learners understand, transform, and authorise its contribution in ways that preserve both their responsibility for the text and a recognisable voice within it.

Acknowledgements

The author would like to thank all participants for their time and valuable contributions to this study.

Disclosure of AI Use

The author used a generative AI tool (ChatGPT) to support language editing, wording refinement, and formatting during manuscript preparation. All research design, data collection, analysis, interpretation, and final decisions about content are the author's own, and the author takes full responsibility for the integrity and content of the manuscript.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Ethics and Consent

Ethical approval was obtained from the university's Permanent Ethics Committee before data collection, and the study was conducted in accordance with the relevant guidelines and regulations. Informed consent was obtained from all participants prior to their participation in the study.

Data Availability

Data are available from the corresponding author on reasonable request.

Appendix A. Survey Instrument (English and Arabic)

The complete survey instrument is reproduced below in the English source version and the Arabic version administered to participants. Reverse-coded items are marked (R) and were recoded prior to analysis. The Arabic is the version administered to participants; it was produced through the translation and independent back-translation procedure described in the Method (Brislin, 1970). Item sources are identified in a note at the end of each section.

Section 1. Background Information

القسم الأول — معلومات عامّة

1.1. Gender

1.1. الجنس

Response options: Male / ذكر; Female / أنثى

1.2. Current year of study

1.2. السنة الدراسية الحالية

Response options: Year 1 / السنة الأولى; Year 2 / السنة الثانية; Year 3 / السنة الثالثة; Year 4 / السنة الرابعة; Year 5 or more / السنة الخامسة أو أكثر

1.3. Self-rated overall English proficiency (CEFR)

1.3. التقييم الذاتي للمستوى العام في اللغة الإنجليزية (الإطار الأوروبي المرجعي)

Response options: A1 Beginner / A1 مبتدئ; A2 Elementary / A2 أساسي; B1 Intermediate / B1 متوسّط; B2 Upper-intermediate / B2 فوق المتوسّط; C1 Advanced / C1 متقدّم; C2 Proficient / C2 متقن

1.4. AI tools used for English writing (multiple selection)

1.4. أدوات الذكاء الاصطناعي المستخدمة في الكتابة بالإنجليزية (اختيار متعدّد)

Response options: ChatGPT, Microsoft Copilot, Google Gemini, Claude, DeepSeek, Grammarly, QuillBot, Other / أخرى

1.5. Duration of AI use for English writing

1.5. مدّة استخدام أدوات الذكاء الاصطناعي في الكتابة بالإنجليزية

Response options: Less than 6 months / أقل من 6 أشهر; 6 months to 1 year / من 6 أشهر إلى سنة; 1 to 2 years / من سنة إلى سنتين; More than 2 years / أكثر من سنتين

Note. Background items were used for sample description. Self-rated proficiency (item 1.3) was coded ordinally from A1 = 1 to C2 = 6 for analysis.

Section 2. AI Use in English Writing

القسم الثاني — استخدام الذكاء الاصطناعي في الكتابة بالإنجليزية

Frequency scale: 1 (Never), 2 (Rarely), 3 (Sometimes), 4 (Often), 5 (Always).

مقياس التكرار: ١ (أبدًا)، ٢ (نادرًا)، ٣ (أحيانًا)، ٤ (غالبًا)، ٥ (دائمًا).

AI1. I use AI tools to generate ideas for my English writing.

AI1. أستخدم أدوات الذكاء الاصطناعي لتوليد الأفكار لكتاباتي بالإنجليزية.

AI2. I use AI tools to check and correct grammar in my English writing.

AI2. أستخدم أدوات الذكاء الاصطناعي لمراجعة وتصحيح القواعد في كتاباتي بالإنجليزية.

AI3. I use AI tools to improve vocabulary choice in my English writing.

AI3. أستخدم أدوات الذكاء الاصطناعي لتحسين اختيار المفردات في كتاباتي بالإنجليزية.

AI4. I use AI tools to paraphrase or reword sentences I have written.

AI4. أستخدم أدوات الذكاء الاصطناعي لإعادة صياغة الجمل التي كتبتها.

AI5. I use AI tools to generate complete paragraphs for me.

AI5. أستخدم أدوات الذكاء الاصطناعي لتوليد فقرات كاملة نيابة عنّي.

AI6. I use AI tools to generate complete essays or texts for me.

AI6. أستخدم أدوات الذكاء الاصطناعي لتوليد مقالات أو نصوص كاملة نيابة عنّي.

AI7. I use AI tools to translate from Arabic into English for my writing.

AI7. أستخدم أدوات الذكاء الاصطناعي للترجمة من العربية إلى الإنجليزية في كتاباتي.

AI8. I use AI tools to make my English writing sound more academic or formal.

AI8. أستخدم أدوات الذكاء الاصطناعي لجعل كتاباتي بالإنجليزية تبدو أكثر أكاديمية أو رسمية.

Note. Items AI1–AI4, AI7, and AI8 formed the supportive AI use composite; items AI5 and AI6 formed the generative AI use composite (full-text generation: complete paragraphs and complete essays). Items were developed for this study, drawing on the use dimension in Xu et al. (2025) and the ChatGPT Usage Scale (Sobaih et al., 2024), and on the purpose-of-use categories in Barrot (2023), Yan (2023), and Xu and Zheng (2025).

Section 3. Perceived Ownership in AI-Assisted Writing

القسم الثالث — الملكية المدركة في الكتابة المدعومة بالذكاء الاصطناعي

Agreement scale: 1 (Strongly Disagree) to 5 (Strongly Agree).

مقياس الموافقة: ١ (لا أوافق بشدّة) إلى ٥ (أوافق بشدّة).

OWN1. When I produce English writing with AI tools, I feel that the writing is mine.

OWN1. عندما أنتج كتابة بالإنجليزية بمساعدة أدوات الذكاء الاصطناعي، أشعر بأنّ هذه الكتابة كتابتي.

OWN2. I feel a high degree of personal ownership over English writing I produce with AI tools.

OWN2. أشعر بدرجة عالية من الملكية الشخصية تجاه الكتابة بالإنجليزية التي أنتجها بمساعدة أدوات الذكاء الاصطناعي.

OWN3. I feel responsible for English writing I produce with AI tools.

OWN3. أشعر بالمسؤولية تجاه الكتابة بالإنجليزية التي أنتجها بمساعدة أدوات الذكاء الاصطناعي.

OWN4 (R). English writing I produce with AI tools does not really feel mine.

OWN4 (R). الكتابة بالإنجليزية التي أنتجها بمساعدة أدوات الذكاء الاصطناعي لا تبدو لي حقًّا كتابتي.

OWN5. I feel a strong personal connection to English writing I produce with AI tools.

OWN5. أشعر بارتباط شخصي قوي بالكتابة بالإنجليزية التي أنتجها بمساعدة أدوات الذكاء الاصطناعي.

OWN6. I feel emotionally connected to English writing I produce with AI tools.

OWN6. أشعر بارتباط عاطفي بالكتابة بالإنجليزية التي أنتجها بمساعدة أدوات الذكاء الاصطناعي.

Note. (R) = reverse-coded. The ownership and responsibility items drew on the felt-ownership and responsibility dimensions of the Psychological Ownership Questionnaire (Avey et al., 2009); the personal-connection and emotional-connection items (OWN5, OWN6) were adapted from Nicholes's (2017) writing-ownership scale. OWN4 is a reverse-coded counterpart included to reduce acquiescence bias.

Section 4. Voice in AI-Assisted Writing

القسم الرابع — الصوت في الكتابة المدعومة بالذكاء الاصطناعي

Agreement scale: 1 (Strongly Disagree) to 5 (Strongly Agree).

مقياس الموافقة: ١ (لا أوافق بشدّة) إلى ٥ (أوافق بشدّة).

VOI1. When I write with AI tools, the final text still sounds like me.

VOI1. عندما أكتب بمساعدة أدوات الذكاء الاصطناعي، لا يزال النص النهائي يشبهني في أسلوبه.

VOI2. English writing I produce with AI tools reflects my personal style.

VOI2. تعكس الكتابة بالإنجليزية التي أنتجها بمساعدة أدوات الذكاء الاصطناعي أسلوبي الشخصي.

VOI3. My own way of expressing ideas comes through in AI-assisted English writing.

VOI3. تظهر طريقتي الخاصة في التعبير عن الأفكار في الكتابة بالإنجليزية المدعومة بالذكاء الاصطناعي.

VOI4 (R). AI-assisted English writing makes my English feel less personal.

VOI4 (R). تجعل الكتابة المدعومة بالذكاء الاصطناعي لغتي الإنجليزية تبدو أقلّ شخصية.

VOI5. I can recognise my own voice in English writing I produce with AI tools.

VOI5. يمكنني التعرّف على صوتي الخاص في الكتابة بالإنجليزية التي أنتجها بمساعدة أدوات الذكاء الاصطناعي.

VOI6 (R). The English I produce with AI tools does not really sound like my own.

VOI6 (R). اللغة الإنجليزية التي أنتجها بمساعدة أدوات الذكاء الاصطناعي لا تبدو حقًّا وكأنّها لغتي أنا.

Note. (R) = reverse-coded. The voice items were newly constructed for this study as a learner self-report measure of authorial voice in AI-assisted L2 writing, for which no established self-report scale existed; they were informed by conceptual accounts of voice in L2 writing rather than adapted from a prior instrument.

References

Alghasab, M. B. (2025). English as a foreign language (EFL) secondary school students' use of artificial intelligence (AI) tools for developing writing skills: Unveiling practices and perceptions. *Cogent Education, 12*(1), Article 2505304. https://doi.org/10.1080/2331186X.2025.2505304

Alharbi, M. A., & Hassan Al-Ahdal, A. A. M. (2025). Exploring Saudi EFL learners' engagement with ChatGPT: A mixed-methods study of perceptions, attitudes, and intentions. *Sage Open, 15*(4). https://doi.org/10.1177/21582440251392080

Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research, 63*(2), 431–463. https://doi.org/10.1177/07356331241307297

Avey, J. B., Avolio, B. J., Crossley, C. D., & Luthans, F. (2009). Psychological ownership: Theoretical extensions, measurement and relation to work outcomes. *Journal of Organizational Behavior, 30*(2), 173–191. https://doi.org/10.1002/job.583

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, Article 100745. https://doi.org/10.1016/j.asw.2023.100745

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Darvin, R. (2025). Identity and investment in the age of generative AI. *Annual Review of Applied Linguistics, 45*, 10–27. https://doi.org/10.1017/S0267190525100135

Draxler, F., Werner, A., Lehmann, F., Hoppe, M., Schmidt, A., Buschek, D., & Welsch, R. (2024). The AI ghostwriter effect: When users do not perceive ownership of AI-generated text but self-declare as authors. *ACM Transactions on Computer-Human Interaction, 31*(2), Article 25. https://doi.org/10.1145/3637875

Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *Language Learning & Technology, 28*(2), 5–31. https://doi.org/10.64152/10125/73570

Helms-Park, R., & Stapleton, P. (2003). Questioning the importance of individualized voice in undergraduate L2 argumentative writing: An empirical study with pedagogical implications. *Journal of Second Language Writing, 12*(3), 245–265. https://doi.org/10.1016/j.jslw.2003.08.001

Hu, H., Zhou, Q., & Hashim, H. (2025). Negotiating identity in the age of ChatGPT: Non-native English researchers' experiences with AI-assisted academic writing. *Humanities and Social Sciences Communications, 12*, Article 965. https://doi.org/10.1057/s41599-025-05351-4

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173–192. https://doi.org/10.1177/1461445605050365

Hyland, K. (2011). Learning to write: Issues in theory, research and pedagogy. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 17–36). John Benjamins. https://doi.org/10.1075/lllt.31.05hyl

Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse*. Cambridge University Press.

Hyland, K. (2026). AI feedback to L2 writers. *International Journal of TESOL Studies, 8*(1), 146–156. https://doi.org/10.58304/ijts.260420

Ivanič, R. (1998). *Writing and identity: The discoursal construction of identity in academic writing*. John Benjamins.

Jacob, S. R., Tate, T., & Warschauer, M. (2025). Emergent AI-assisted discourse: A case study of a second language writer authoring with ChatGPT. *Journal of China Computer-Assisted Language Learning, 5*(1), 1–22. https://doi.org/10.1515/jccall-2024-0011

Jiang, F. K., & Hyland, K. (2025). Does ChatGPT write like a student? Engagement markers in argumentative essays. *Written Communication, 42*(3), 463–492. https://doi.org/10.1177/07410883251328311

Joshi, N., & Vogel, D. (2025). Writing with AI lowers psychological ownership, but longer prompts can help. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)* (Article 72, pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3719160.3736608

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv*. https://doi.org/10.48550/arXiv.2506.08872

Matsuda, P. K. (2001). Voice in Japanese written discourse: Implications for second language writing. *Journal of Second Language Writing, 10*(1–2), 35–53. https://doi.org/10.1016/S1060-3743(00)00036-9

Mizumoto, A., Yasuda, S., & Tamura, Y. (2024). Identifying ChatGPT-generated texts in EFL students' writing: Through comparative analysis of linguistic fingerprints. *Applied Corpus Linguistics, 4*(3), Article 100106. https://doi.org/10.1016/j.acorp.2024.100106

Nañola, E. L., Arroyo, R. L., Hermosura, N. J. T., Ragil, M., Sabanal, J. N. U., & Mendoza, H. B. (2025). Recognizing the artificial: A comparative voice analysis of AI-generated and L2 undergraduate student-authored academic essays. *System, 130*, Article 103611. https://doi.org/10.1016/j.system.2025.103611

Nicholes, J. (2017). Measuring ownership of creative versus academic writing: Implications for interdisciplinary praxis. *Writing in Practice, 3*. https://doi.org/10.62959/WIP-03-2017-08

Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). SAGE Publications.

Pierce, J. L., Kostova, T., & Dirks, K. T. (2003). The state of psychological ownership: Integrating and extending a century of research. *Review of General Psychology, 7*(1), 84–107. https://doi.org/10.1037/1089-2680.7.1.84

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Sandstead, M., & Kibler, A. (2025). Voice in L2 writing in the age of AI. *Journal of Second Language Writing, 69*, Article 101212. https://doi.org/10.1016/j.jslw.2025.101212

Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2025). Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2025.2454541

Sobaih, A. E. E., Elshaer, I. A., & Hasanein, A. M. (2024). Examining students' acceptance and use of ChatGPT in Saudi Arabian higher education. *European Journal of Investigation in Health, Psychology and Education, 14*(3), 709–721. https://doi.org/10.3390/ejihpe14030047

Tardy, C. M. (2012). Current conceptions of voice. In K. Hyland & C. Sancho Guinda (Eds.), *Stance and voice in written academic genres* (pp. 34–48). Palgrave Macmillan.

Teng, M. F. (2024a). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*, Article 100270. https://doi.org/10.1016/j.caeai.2024.100270

Teng, M. F. (2024b). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies, 6*(3), 36–57. https://doi.org/10.58304/ijts.20240304

Teng, M. F. (2025a). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education, 60*(1), Article e12811. https://doi.org/10.1111/ejed.12811

Teng, M. F. (2025b). Understanding EFL student writers' metacognitive awareness in utilizing ChatGPT. *System, 135*, Article 103848. https://doi.org/10.1016/j.system.2025.103848

Teng, M. F., & Huang, J. (2026). Assessing ChatGPT feedback for EFL learners' engagement and writing performance. *International Journal of Applied Linguistics*. Advance online publication. https://doi.org/10.1111/ijal.70155

Teng, M. F., & Shen, X. (2025). Metacognitive awareness and EFL learners' writing performance in ChatGPT-assisted writing context. *International Journal of Applied Linguistics*. Advance online publication. https://doi.org/10.1111/ijal.70049

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing, 67*, Article 101187. https://doi.org/10.1016/j.jslw.2025.101187

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for second language writers. *Journal of Second Language Writing, 62*, Article 101071. https://doi.org/10.1016/j.jslw.2023.101071

Xu, J., & Zheng, Y. (2025). Negotiating understanding, control, and authorship: L2 learners' experiences with AI-assisted paraphrasing in academic writing. *Journal of English for Academic Purposes, 78*, Article 101591. https://doi.org/10.1016/j.jeap.2025.101591

Xu, L., Zhang, L., Ou, L., & Wang, D. (2025). The development and validation of a scale on student AI literacy in L2 writing: A domain-specific perspective. *Journal of Second Language Writing, 69*, Article 101227. https://doi.org/10.1016/j.jslw.2025.101227

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*(11), 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

Yoon, H.-J. (2017). Textual voice elements and voice strength in EFL argumentative writing. *Assessing Writing, 32*, 72–84. https://doi.org/10.1016/j.asw.2017.02.002

Zhao, C.-G. (2013). Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric. *Language Testing, 30*(2), 201–230. https://doi.org/10.1177/0265532212456965

Zhao, C.-G., & Wu, J. (2022). Perceptions of authorial voice: Why discrepancies exist. *Assessing Writing, 53*, Article 100632. https://doi.org/10.1016/j.asw.2022.100632

Aser Altalib (PhD, Australian National University) is an Assistant Professor of Applied Linguistics at Jouf University, Saudi Arabia. His research focuses on affective, motivational, and psychological factors in second language learning and teaching, including L2 motivation, enjoyment, anxiety, burnout, self-efficacy, emotional intelligence, teacher self-efficacy, and the role of generative AI in English language education.

ORCID: https://orcid.org/0000-0003-1555-3188