



Generative AI Use and Authorial Voice among Pakistani University ESL Learners

Muhammad Aslam Khan^{*a}, Dr. Muhammad Akram^b

a. PhD Scholar in the Department of English Linguistics, The Islamia University of Bahawalpur, Pakistan

b. Assistant Professor, Department of English Linguistics, The Islamia University of Bahawalpur, Pakistan

ARTICLE INFO

Received:

July 07, 2026

Revision Received:

July 24, 2026

Accepted:

August 12, 2026

Available Online:

August 27, 2026

Keywords:

Generative artificial intelligence; authorial voice; ESL academic writing; writer agency; AI literacy; mixed methods; Pakistan

ABSTRACT

Generative artificial intelligence (GenAI) is rapidly transforming second-language academic writing, but its association with authorship and authorial voice remains insufficiently understood. This explanatory sequential mixed-methods study examined whether editorial use, dialogic use, and generative delegation were differentially associated with authorial voice among 92 university ESL learners in Southern Punjab, Pakistan. Quantitative evidence comprised survey measures, one argumentative essay per participant, corpus-based linguistic indicators, and independent ratings using an analytic voice rubric; qualitative explanation was provided by interviews with 14 purposively selected participants. Generative delegation was negatively correlated with self-perceived authorial voice ($r = -.45, p < .001$) and rubric-scored textual voice ($r = -.44, p < .001$), whereas dialogic use was positively correlated with self-perceived voice and editorial use was not significantly related to either voice outcome. The final regression model accounted for 52.0% of the variance in self-perceived voice; writer agency had the largest standardized coefficient. The reported mediation analysis indicated that agency accounted for 43.6% of the association between delegation and self-perceived voice, whereas the delegation-by-AI-literacy interaction was not statistically significant. Greater delegation was also associated with fewer self-mentions and hedges, but with more boosters, longer sentences, and greater lexical diversity. Interviews generated four themes: outsourcing the difficult middle, the fluency-ownership trade-off, the disappearing draft, and voice as risk management. They nevertheless suggest that pedagogical guidance should distinguish assistance that supports writers' decisions from delegation that replaces core composing decisions.

1. Introduction

Within three years of the public release of ChatGPT, generative artificial intelligence has moved from novelty to infrastructure in university writing. Surveys across very different higher-education systems now report majority adoption among students, with the tools used at every stage of composition from idea generation to final proofreading (Almassaad et al., 2024; Abbas et al., 2024). In second-language contexts, the uptake has been even faster, because the tools appear to solve precisely the problem that has historically consumed most of an ESL writer's attention: producing grammatically acceptable, idiomatically plausible academic English (Barrot, 2023; Warschauer et al., 2023). Early empirical work was accordingly framed around benefits. Studies have reported gains in grammatical accuracy, lexical range, cohesion, and organization, together with reductions in writing anxiety and increases in engagement (Su et al., 2023; Yan, 2023; Song & Song, 2023; Hongxia & Razali, 2025). A second

* Correspondence to: PhD Scholar in the Department of English Linguistics, The Islamia University of Bahawalpur, Pakistan
E-mail address: aslambaloch33@gmail.com (M.A.Khan)



literature has grown alongside the first and asks a different question. If the fluent surface of a student text no longer indexes the student's own developing control of English, what exactly is being assessed, and what is being learned? Concerns about dependency, integrity, and the concealment of weak reasoning behind polished language have been raised repeatedly (Perkins, 2023; Kasneci et al., 2023; Lee et al., 2025). Among these concerns, one is peculiarly resistant to the usual remedies of detection software and honor codes, because it does not involve deception at all. A student may declare their use of GenAI, comply fully with institutional policy, and still hand in a text in which nothing sounds like them. This is the problem of authorial voice.

1.1. Statement of the Problem

Authorial voice has been theorized in applied linguistics for three decades as the discursively constructed presence of a writer in a text, realized through self-mention, stance, evaluation, hedging, and the management of alignment with a reader (Ivanic, 1998; Hyland, 2005, 2008; Matsuda, 2001; Tardy, 2012). It is not decoration. Voice is how an academic writer signals what they are committed to, how strongly, and against which alternatives, and it is one of the things that distinguishes a novice text that reports from an expert text that argues. For multilingual learners, developing voice is also identity work, since it involves finding a position from which to speak inside a discourse community that was not built around them (Ivanic & Camps, 2001; Canagarajah, 2015). Recent comparative studies indicate that large language models produce text with a systematically different voice profile from student text: more assertive, more collective or more impersonally expert, less locally evaluative, and stripped of the small hesitations and idiosyncrasies that mark a developing writer (Amirjalili et al., 2024; Al Hosni, 2025). If ESL learners increasingly submit text that has passed through such a model, the voice profile of their writing may converge on the model's rather than developing towards their own. What is missing from the literature is empirical evidence, at the level of the individual writer, connecting how a student uses GenAI to what happens to voice in the text they submit and to how they experience their own authorship. Almost all existing evidence is either textual, comparing AI essays with human essays produced independently, or attitudinal, surveying perceptions without examining any writing. The two have rarely been joined.

1.2. Purpose and Objectives

This study joins them. It treats GenAI use not as a single quantity but as three functionally distinct modes: editorial use, in which the tool corrects and polishes text the writer has already produced; dialogic use, in which the tool is used to think with, for brainstorming, counterargument, feedback, or explanation; and generative delegation, in which the tool produces the prose that will be submitted, with the writer editing afterwards. The objectives were: to establish which modes of GenAI use predict self-perceived authorial voice and rubric-measured textual voice among Pakistani university ESL learners; to test whether writer agency, that is, the sense of owning and controlling one's own text, mediates the relationship between generative delegation and voice; to test whether AI literacy moderates, and thereby buffers, that relationship; to describe how the linguistic markers conventionally used to operationalize voice differ across levels of delegation and to explain the quantitative pattern through participants' own accounts of how they write with these tools.

1.3. Research Questions

Five research questions were addressed. RQ1: How intensively, and in which modes, do Pakistani university ESL learners use GenAI in academic writing? RQ2: How are editorial use, dialogic use, and generative delegation related to self-perceived authorial voice, textual voice intensity, perceived voice loss, and writer agency? RQ3: Does writer agency statistically mediate the association between generative delegation and self-perceived authorial voice? RQ4: Does AI literacy moderate the association between generative delegation and self-perceived authorial voice? RQ5: How do learners explain changes in the ownership and perceived sound of their writing when using GenAI?

2. Literature Review

2.1. Conceptualizing Authorial Voice

Voice entered composition studies as an expressivist idea, the authentic individual sounding through the prose (Elbow, 1994). Applied linguistics recast it as social and discursive. Ivanic (1998) distinguished the autobiographical self, the discursive self-constructed in the text, and the self as author, the degree to which a writer claims authority for their own propositions. Hyland (2005, 2008) supplied the analytic apparatus, describing voice as the interaction of stance, comprising hedges, boosters, attitude markers, and self-mention, with engagement, the moves by which a writer positions a reader. Matsuda (2001) reframed voice as the reader's impression created by an accumulation of discursive choices, which made it measurable without recourse to authorial intention. Zhao (2013) operationalized it for L2 assessment in an analytic rubric with three dimensions: presence and clarity of ideas, strength of the authorial voice in the argument, and the writer-reader relationship. Two features of this tradition matter for the present study. First, voice is realized through countable surface features, which means that a change in voice leaves textual traces. Second, voice for L2 learners is developmental and effortful. It is acquired late, it costs more cognitively than propositional accuracy, and it is the first thing to disappear under pressure (Zhao & Llosa, 2008; Stapleton, 2002). Anything that removes the pressure by removing the composing itself is therefore likely to remove voice too.

2.2. Generative AI in L2 Writing

The empirical literature on GenAI in L2 writing has expanded rapidly and is now large enough to support syntheses (Frontiers synthesis of 2023 to 2025 evidence; Al Hosni, 2025). The consistent findings are that GenAI improves the formal properties of student texts, that it functions well as an on-demand feedback provider, and that students value it most for the very features they find hardest, namely lexical choice, cohesion, and organization (Chen et al., 2024; Dong, 2025; Alkamel & Alwagieh, 2024). Experimental and quasi-experimental designs report gains in writing performance and self-efficacy (Kiziltas, 2025; Franjic & Bozinovski, 2025), although effects vary with proficiency, with lower-proficiency learners generally benefiting more. Counter-evidence concerns the process rather than the product. Studies on how students work with these tools find rapid uptake of suggestions without evaluation, reduced revision effort, and self-reported declines in cognitive effort (Levine et al., 2025; Lee et al., 2025). Learners themselves report awareness of over-reliance and an intention to reduce it, which suggests that the problem is perceived from inside as well as diagnosed from outside. Alzubi et al. (2025) found that generative tools can either suppress or nurture creativity depending on how their use is structured, which points towards mode of use rather than amount of use as the operative variable, a position the present study adopts directly.

2.3. Voice, Authorship and Machine Text

A smaller body of literature compares the voice profiles of AI-generated and human-authored academic texts. Amirjalili et al. (2024) compared a student essay with a ChatGPT essay on the same prompt and found the machine text formally competent but weak in self-identification and authorial presence. Comparative rubric-based analyses of larger essay sets report that student essays are dominated by one voice type while AI generators favor another, and that AI essays exhibit an elusive textual identity that resists attribution. Al Hosni's (2025) thematic review of eighteen studies concluded that AI-generated academic prose is grammatically polished and structurally coherent but lacks emotional nuance, rhetorical subtlety, and the distinctive markers that constitute a linguistic fingerprint, and that this poses a particular threat in ESL settings. Two limitations run through this work. It compares texts rather than learners, so it cannot say what happens to a given writer's voice as their use changes; and it typically compares fully AI-generated with fully human text, whereas real student writing is a hybrid produced by iterative prompting and editing. The present study addresses both by measuring use, self-perception, and text within the same individuals.

2.4. Writer Agency and Textual Ownership

Agency in writing research denotes the writer's experienced capacity to make and own compositional choices. It is closely tied to authorship beliefs and to the willingness to take a position, and it is the psychological correlate of the self-as-author dimension of voice (Ivanic, 1998). When a text is produced by another party and then adopted, ownership becomes ambiguous even when authorship is formally retained, a condition described in the GenAI literature as authorial ambiguity and as a source of cognitive dissonance for the writer (Seran et al., 2025; Ryan et al., 2025). Duah-type accounts of blurred authorial identity converge on the same point. Agency is therefore the natural candidate mediator: delegation does not act on voice directly, but rather dissolves the sense of authorship from which voice is written.

2.5. AI Literacy as a Possible Buffer

If uncritical adoption is the mechanism of harm, then critical evaluative capacity should be the remedy. AI literacy, understood as the ability to judge, verify, and selectively resist model output, has been proposed as the central pedagogical response (UNESCO, 2023; Kohnke et al., 2023). Alghamdi (2025) found that self-efficacy and disciplinary background shaped whether students perceived GenAI as supporting or replacing their thinking. Whether AI literacy actually attenuates the effect of delegation on voice, as opposed to simply raising voice on average, is an open empirical question, and this study tests it as a moderation hypothesis.

2.6. The Pakistani Higher-Education Context

To date, Pakistani studies have been dominated by perception surveys and technology acceptance models. Rashid et al. (2025) documented the knowledge, use, and perceived impact of AI on academic writing among Pakistani students, reporting broad use alongside limited critical training. Anwer et al. (2025) found strong positive associations between trust in ChatGPT and perceived improvement in clarity, vocabulary, and organization in a Pakistani ESL sample. Alghazo et al. (2025) examined the integrity and ethical challenges from Pakistani students' perspectives. However, none of these studies examined students' writing. The infrastructural picture matters too: uneven connectivity, large classes, limited writing center provision, and assessment regimes that reward formal correctness combine to make delegation both attractive and low risk.

3. Methodology

3.1. Research Design

An explanatory sequential mixed-methods design was used (Creswell & Plano Clark, 2018). Phase one combined a cross-sectional survey with a text-analytic study of the same participants' argumentative essays. Phase two used semi-structured interviews with a purposive subsample selected from the extremes of the delegation distribution to explain the quantitative results rather than merely supplement them. Priority was given to the quantitative results, with qualitative data used for explanation and integration at the interpretation stage through a joint display.

3.2. Setting and Participants

The participants were 92 ESL learners enrolled in English, social science, and STEM programmes at three public-sector universities in Southern Punjab, Pakistan. All were second-language users of English studying through English-medium instruction at university. Convenience sampling with quota targets for level of study yielded 44 men and 48 women: 41 BS, 32 MPhil/MS, and 19 PhD students; 37 studied humanities/English, 28 social sciences, and 27 STEM disciplines. Mean age was 23.19 years ($SD = 3.10$; range = 18-32), and mean duration of formal English instruction was 13.66 years ($SD = 2.33$). Fifty-one participants had attended English-medium schools and 41 had attended Urdu- or regional-language-medium schools. Inclusion required at least one semester of coursework involving written assignments and self-reported use of at least one GenAI tool during the preceding six months. Of 108 students approached, 92 provided usable survey and essay data (response rate = 85.2%). A sensitivity analysis for a regression with 10 predictors, $\alpha = .05$, $N = 92$, and 80% power indicated a minimum detectable effect of approximately $f^2 = .20$; therefore, smaller unique effects may have gone undetected.

3.3. Instruments

A composite questionnaire was assembled. Part A collected demographic and background information, including tools used, frequency of use, and assignment types. Part B contained eight scales rated from 1 (strongly disagree) to 5 (strongly agree): Editorial Use (6 items), Dialogic Use (6 items), Generative Delegation (7 items), AI Literacy (8 items), Writing Self-Efficacy (8 items), Writer Agency (8 items), Self-Perceived Authorial Voice (10 items), and Perceived Voice Loss (6 items). Writing Self-Efficacy was developed in accordance with Bandura's (2006) scale-construction guidance, and the authorial-voice items were aligned with Zhao's (2013) rubric dimensions. The three use-mode scales and Perceived Voice Loss were developed for this study because an established instrument separating GenAI use by function was not identified. Three applied linguists reviewed the item pool, and pilot testing informed revision; item-level content validity indices exceeded .80 for all retained items. Complete items, scoring instructions, pilot-sample characteristics, and evidence for the scales' dimensionality should be supplied in an appendix or supplementary file before submission.

3.4. Essay Corpus and Voice Rubric

Each participant submitted one argumentative essay of 700-1,000 words written for a course assignment during the same semester and granted permission for its analysis. Two trained raters independently scored each essay using an analytic voice rubric adapted from Zhao (2013), producing a Textual Voice Intensity score from 1 to 4 across three dimensions. Inter-rater reliability for the composite was $ICC = .84$ using a two-way random-effects model; the manuscript should specify whether absolute agreement or consistency and whether the reported coefficient refers to single or average measures. Disagreements greater than one scale point were resolved through discussion after the independent ratings. Each essay was also analyzed for six linguistic measures: first-person self-mentions, hedges, boosters, and attitude markers per 1,000 words; type-token ratio; and mean sentence length. Marker lists followed Hyland (2005), and every automated hit was manually checked to exclude non-stance uses. The tokenizer, software/version, lexical-diversity preprocessing, and full marker lists should be reported in a reproducible supplement before submission.

3.5. Interview Protocol

Fourteen participants were interviewed; seven were drawn from the top and seven from the bottom tertile of the delegation distribution, with levels of study and gender balanced as far as possible. Interviews lasted 32 to 55 minutes, were conducted in a mixture of English and Urdu according to participant preference, audio-recorded with consent, transcribed and, where necessary, translated with back-translation checks. The protocol covered a walkthrough of the most recent assignment, points at which the tool was consulted, the experience of editing model output, feelings of ownership, and perceptions of how their writing sounded.

3.6. Procedure

Data were collected during one academic semester. After departmental permission, potential participants received an information sheet and completed electronic informed consent before accessing the online questionnaire. Participants uploaded

their own essays through the same form, and survey and essay data were linked using anonymized codes. Preliminary survey analysis was completed before interview sampling; interviews were conducted approximately four weeks later so that participants could be selected from the upper and lower tertiles of the delegation distribution.

3.7. Data Analysis

Quantitative analysis comprised data screening, internal-consistency estimation, descriptive statistics, zero-order Pearson correlations, independent-samples *t* tests, one-way ANOVA with Tukey HSD where applicable, three-block hierarchical multiple regression, bootstrap mediation with 5,000 resamples, and moderated regression using mean-centered predictors and an interaction term (Preacher & Hayes, 2008; Hayes, 2022). Regression diagnostics included variance inflation factors, the Durbin-Watson statistic, and visual inspection of residual plots. Harman's single-factor test was reported as a limited diagnostic for common-method variance (Podsakoff et al., 2003), but it does not rule out common-method bias. The authors should report the statistical software and version, missing-data handling, outlier criteria, assumption-test results, exact confidence-interval method, coding of categorical predictors, and complete coefficients with 95% confidence intervals. Interview data were analyzed using a structured codebook thematic approach: two coders double-coded 20% of transcripts, discussed disagreements, and refined the codebook ($\kappa = .81$). Quantitative and qualitative findings were integrated at interpretation through comparison of convergence, complementarity, and divergence.

Table 1: Demographic and Background Characteristics of Participants

Characteristic	Category	n	%
Gender	Male	44	47.8
	Female	48	52.2
Level of study	BS (undergraduate)	41	44.6
	MPhil / MS	32	34.8
	PhD	19	20.7
Discipline	Humanities / English	37	40.2
	Social sciences	28	30.4
	STEM	27	29.3
Medium of prior schooling	English medium	51	55.4
	Urdu / regional medium	41	44.6
Age (years)	M = 23.19, SD = 3.10	-	-
Years of English instruction	M = 13.66, SD = 2.33	-	-

Note. Percentages are rounded and may not total 100. Age range: 18 to 32 years.

Table 2: Instruments, Item Numbers and Internal Consistency

Scale	k	Cronbach's α
Editorial use	6	.859
Dialogic use	6	.861
Generative delegation	7	.894
AI literacy	8	.903
Writing self-efficacy	8	.885
Writer agency	8	.899
Authorial voice (self-perceived)	10	.914
Perceived voice loss	6	.876

Note. *k* = number of items. All items rated 1 to 5. Textual Voice Intensity was rater-scored, not a self-report scale; its inter-rater ICC was .84.

4. Results

4.1. Participant Characteristics

All 92 cases contributed survey and essay data. The sample comprised 44 men and 48 women across BS, MPhil/MS, and PhD levels; full demographic characteristics are shown in Table 1. The manuscript should state explicitly whether item-level data were complete and, if not, how missing responses were handled.

4.2. Reliability and Descriptive Statistics

Internal consistency was good to excellent for the eight self-report scales, with Cronbach's alpha ranging from .859 to .914 (Table 2). Descriptive statistics are presented in Table 3. Editorial use was the most frequently reported mode ($M = 3.92$, $SD = 0.58$), followed by dialogic use ($M = 3.28$, $SD = 0.70$) and generative delegation ($M = 3.05$, $SD = 0.82$). Forty participants (43.5%) scored at or above 3.5 on the delegation scale, indicating relatively frequent endorsement of delegation items; this threshold was descriptive and was not a validated clinical or behavioral cut-off. Self-perceived authorial voice ($M = 3.29$, $SD = 0.63$) and writer agency ($M = 3.36$, $SD = 0.66$) were near the midpoint-to-upper range. Rubric-scored textual voice averaged 2.61 ($SD = 0.58$) on the four-point scale, and perceived voice loss averaged 3.12 ($SD = 0.74$). Interpretive labels such as low, moderate, or high should be used only if prespecified and validated cut-points are available.

Table 3: Descriptive Statistics for Study Variables

Variable	M	SD	Min	Max	Skewness	Kurtosis
Editorial use	3.92	0.58	2.25	5.00	-.05	.07
Dialogic use	3.28	0.70	1.48	5.00	.20	-.11
Generative delegation	3.05	0.82	1.00	5.00	-.14	-.28
AI literacy	3.31	0.66	1.81	4.83	.08	-.26
Writing self-efficacy	3.41	0.63	1.71	4.85	-.18	-.15
Writer agency	3.36	0.66	1.93	4.92	.09	-.44
Authorial voice (self-perceived)	3.29	0.63	1.68	5.00	.14	.56
Perceived voice loss	3.12	0.74	1.45	4.76	-.05	-.69
Textual voice intensity (1-4)	2.61	0.58	1.17	3.91	-.23	-.19

Note. All scales scored 1 to 5, except textual voice intensity, which was rater-scored 1 to 4.

Table 4: Linguistic Voice Markers in the Essay Corpus

Measure	M	SD	Min	Max
First-person self-mention / 1000 words	9.83	3.26	0.50	16.81
Hedges / 1000 words	14.40	3.65	5.01	23.25
Boosters / 1000 words	8.09	3.11	1.52	15.05
Attitude markers / 1000 words	6.22	2.74	0.50	14.20
Type-token ratio	0.412	0.021	0.366	0.451
Mean sentence length (words)	21.70	2.95	14.97	28.64

Note. The stance categories follow Hyland (2005). All automatic hits were manually verified.

Reported tool use was concentrated: ChatGPT was used by 78.3 percent of participants, QuillBot by 55.4 percent, Grammarly by 47.8 percent, Gemini by 39.1 percent, Copilot by 21.7 percent, and DeepSeek by 18.5 percent, with most participants using more than one. Corpus measures across the 92 essays (Table 4) were: self-mention 9.83 per 1000 words ($SD = 3.26$), hedges 14.40 ($SD = 3.65$), boosters 8.09 ($SD = 3.11$), attitude markers 6.22 ($SD = 2.74$), type-token ratio .412 ($SD = .021$) and mean sentence length 21.70 words ($SD = 2.95$).

4.3. Correlations

Zero-order correlations are reported in Table 5. Editorial use was not significantly correlated with self-perceived voice ($r = -.02$, $p = .843$), textual voice ($r = -.08$, $p = .440$), or agency ($r = -.07$, $p = .500$). Dialogic use was positively correlated with self-perceived voice ($r = .39$, $p < .001$), textual voice ($r = .23$, $p = .029$), agency ($r = .25$, $p = .016$), and AI literacy ($r = .37$, $p < .001$). Generative delegation was negatively correlated with self-perceived voice ($r = -.45$, $p < .001$), textual voice ($r = -.44$, $p < .001$), agency ($r = -.40$, $p < .001$), and AI literacy ($r = -.22$, $p = .034$), and positively correlated with perceived voice loss ($r = .48$, $p < .001$). These bivariate associations address RQ2 but do not establish directional effects.

Table 5: Zero-Order Pearson Correlations Among Study Variables

Variable	1	2	3	4	5	6	7	8	9
1. Editorial use	-								
2. Dialogic use	.02	-							
3. Delegation	.24*	-.16	-						
4. AI literacy	-.04	.37***	-.22*	-					
5. Self-efficacy	.02	.31**	-.15	.20	-				
6. Writer agency	-.07	.25*	-.40***	.25*	.40***	-			

7. Voice (self)	-.02	.39***	-.45***	.44***	.40***	.56***	-		
8. Voice loss	.14	-.07	.48***	-.03	-.16	-.46***	-.38***	-	
9. Textual voice	-.08	.23*	-.44***	.21*	.30**	.49***	.69***	-.29**	-

Note. * $p < .05$. ** $p < .01$. *** $p < .001$. Delegation also correlated with self-mention ($r = -.76$), hedges ($r = -.58$), and boosters ($r = .55$), all $p < .001$.

Delegation was also correlated with self-mention ($r = -.76$), hedges ($r = -.58$), and boosters ($r = .55$), all $p < .001$. These strong associations indicate covariation between reported delegation and selected linguistic features at the sample level; they should not be interpreted as prediction, causation, or an individual-level AI-detection rule. Editorial use was modestly correlated with delegation ($r = .24$, $p = .021$), suggesting that the two behaviors can co-occur while remaining empirically distinct.

4.4. Group differences

No statistically significant gender difference was detected for any listed variable (all $p > .51$; $|d| \leq 0.14$; Table 6). Participants with English-medium schooling reported higher AI literacy than those with Urdu- or regional-language-medium schooling ($M = 3.45$ vs. 3.14), $t(90) = -2.29$, $p = .024$, $d = -0.48$; the sign reflects the group-order coding in Table 6. The writing self-efficacy comparison was not statistically significant ($p = .078$). Neither delegation nor either voice measure differed significantly by prior medium of schooling. Because multiple unadjusted comparisons were performed, the isolated $p = .024$ result should be interpreted cautiously and accompanied by confidence intervals in the final table.

Table 6: Independent-Samples t Tests by Gender and by Medium of Prior Schooling

Variable	Male M	Female M	t	d	Urdu M	English M	t	d
Editorial use	3.91	3.93	-0.14	-.03	3.94	3.91	0.21	.04
Dialogic use	3.28	3.28	0.02	.01	3.23	3.32	-0.62	-.13
Generative delegation	3.11	3.00	0.65	.14	3.08	3.03	0.30	.06
AI literacy	3.29	3.32	-0.23	-.05	3.14	3.45	-2.29*	-.48
Writing self-efficacy	3.43	3.39	0.27	.06	3.28	3.51	-1.79	-.37
Writer agency	3.33	3.38	-0.36	-.08	3.31	3.40	-0.68	-.14
Authorial voice	3.26	3.31	-0.44	-.09	3.23	3.33	-0.80	-.17
Perceived voice loss	3.08	3.15	-0.45	-.09	3.13	3.11	0.08	.02
Textual voice	2.60	2.62	-0.16	-.03	2.66	2.57	0.74	.16

Note. $df = 90$ for all tests. $d =$ Cohen's d . * $p < .05$. No significant gender comparison was found. AI literacy differed by medium of prior schooling, $t(90) = -2.29$, $p = .024$.

4.5. Predicting Self-Perceived Authorial Voice

Table 7: Hierarchical Multiple Regression Predicting Self-Perceived Authorial Voice

Predictor (standardized beta)	Model 1	Model 2	Model 3
Gender	.08	.04	.03
Level of study	.23*	.17	.17
Years of English	.07	.03	-.05
Medium of schooling	.08	.06	-.00
Editorial use	-	.06	.05
Dialogic use	-	.30**	.14
Generative delegation	-	-.41***	-.24**
AI literacy	-	-	.15
Writing self-efficacy	-	-	.16
Writer agency	-	-	.35***
R ²	.067	.345	.520
Adjusted R ²	.024	.290	.461
Delta R ²	-	.277***	.175***
F	1.57	6.31***	8.77***

Note. * $p < .05$. ** $p < .01$. *** $p < .001$. Model 1 $df = (4, 87)$; Model 2 $df = (7, 84)$; Model 3 $df = (10, 81)$. Delta R² for Model 2, $F(3, 84) = 11.86$, $p < .001$; for Model 3, $F(3, 81) = 9.86$, $p < .001$. VIF ranged from 1.07 to 1.42, and Durbin-Watson was 2.16.

Hierarchical regression results appear in Table 7. Block 1, containing demographic controls, was not significant, $R^2 = .067$, $F(4, 87) = 1.57$, $p = .191$. Adding the three use modes in Block 2 increased explained variance by Delta R² = .278, F -change(3, 84) =

11.86, $p < .001$; delegation ($\beta = -.406$, $p < .001$) and dialogic use ($\beta = .298$, $p = .001$) were significant, whereas editorial use was not ($\beta = .057$, $p = .534$). Adding AI literacy, self-efficacy, and agency in Block 3 increased explained variance by Delta $R^2 = .175$, $F\text{-change}(3, 81) = 9.86$, $p < .001$. The final model explained 52.0% of the variance, $F(10, 81) = 8.77$, $p < .001$, adjusted $R^2 = .461$. In the final model, writer agency had the largest standardized coefficient ($\beta = .350$, $p < .001$), and delegation remained negatively associated with self-perceived voice ($\beta = -.241$, $p = .007$). Dialogic use ($\beta = .139$, $p = .111$), AI literacy ($\beta = .148$, $p = .128$), and self-efficacy ($\beta = .161$, $p = .078$) did not make statistically significant unique contributions. Their attenuation does not, by itself, demonstrate mediation. Unstandardized coefficients, standard errors, 95% confidence intervals, predictor coding, and the full assumption checks should be added to Table 7 before submission.

4.6. Mediation and Moderation Analyses

The manuscript reports that writer agency statistically accounted for 43.6% of the association between generative delegation and self-perceived authorial voice and that the delegation-by-AI-literacy interaction was not statistically significant. However, the present file does not provide the a, b, c, and c-prime paths, bootstrap indirect effect and 95% confidence interval, total effect, interaction coefficient, standard error, confidence interval, change in R^2 , or conditional effects. These values must be copied from the verified analysis output and reported before the mediation and moderation conclusions can be independently evaluated. No numerical values should be reconstructed or invented.

4.7. Corpus Associations

The corpus results were association-based. Delegation correlated negatively with self-mention and hedging and positively with boosting (Table 5 note). If group-based claims about longer sentences or greater lexical diversity are retained in the abstract, discussion, or conclusion, the corresponding coefficients or prespecified group comparisons, uncertainty intervals, and p values must be added to the Results section.

4.8. Qualitative Findings

Four themes were developed from the 14 interviews. The manuscript should add a compact joint display reporting each theme, its analytic meaning, the number of participants contributing to it where frequency is relevant, one anonymized illustrative quotation, and the quantitative result that it helps explain. Participant counts should not be treated as prevalence estimates in this purposively selected qualitative subsample.

4.8.1. Outsourcing the difficult middle

High-delegation participants described a consistent division of labor. They chose the topic, knew broadly what they wanted to argue, and edited the result. They handed over the middle: turning a position into paragraphs. One MPhil participant described knowing exactly what she wanted to say and using the tool because getting it into acceptable English took hours, she did not have. This is significant theoretically, because the stage being outsourced is Flower and Hayes's (1981) translation stage, where stance is encoded. Participants did not experience this as cheating, and several were explicit that the ideas remained theirs. The loss they were unaware of was not of content but of the encoding of position.

4.8.2. The fluency-ownership trade-off

Participants across both tertiles articulated a trade-off in almost identical terms: the text improved and their claim on it weakened. Several described reading their submitted work back and finding it correct but foreign, or recognizing the argument as theirs while the sentences were not. Low-delegation participants described the inverse, defending prose they knew to be clumsier because it was theirs. The trade was consciously made, and, importantly, most participants judged it worth making under assessment conditions that reward correctness.

4.8.3. The disappearing draft

Nine of the fourteen participants described a striking structural change: the messy first draft stopped existing. Writing once began with a rough attempt that was then revised; it now begins with fluent generated prose that is edited down. Participants described losing the experience of not knowing what they thought until they had written it poorly first. Since voice in the L2 development literature emerges through the iterative revision of one's own inadequate text, the disappearance of the inadequate text removes the site where the voice was learned. This theme explains why the delegation effect is unrelated to the level of study: the mechanism is procedural, not developmental, and applies equally to a doctoral candidate.

4.8.4. Voice as risk management

The fourth theme was unexpected and inverts the standard framing. Several participants described deliberately removing personal markers from their writing because detection tools, supervisors, and departmental conventions treat first-person expression and hedged qualification as unprofessional or symptoms of AI use. One PhD participant reported being told to

remove first-person constructions entirely; another said that anything that sounds like a person invites suspicion. Voice suppression here is a rational response to an assessment environment, not a side effect of tool use. Any intervention that ignores this will fail, because it will ask students to take a risk that their institutions currently punish.

5. Discussion

The headline finding is that the three modes of GenAI use behave as three different variables. Editorial use showed no relationship with any voice outcome, dialogic use showed positive relationships, and generative delegation showed strong negative relationships with both self-perceived and rubric-measured voice. Studies that treat GenAI use as a single quantity are therefore averaging over effects that run in opposite directions, which may explain some of the inconsistency in the existing literature, where positive findings for writing quality (Su et al., 2023; Yan, 2023) sit alongside concerns about dependency and authorship (Perkins, 2023; Lee et al., 2025). Both can be true because they describe different behaviors. This finding aligns with Alzubi et al. (2025), who found that generative tools either suppress or nurture creativity depending on how their use is structured, and convert that observation into a measurement recommendation: use should be measured functionally.

The reported mediation result suggests that writer agency may help explain the association between delegation and self-perceived voice, which is consistent with Ivanic's (1998) self-as-author construct and accounts of authorial ambiguity (Seran et al., 2025; Ryan et al., 2025). This interpretation must remain statistical rather than causal because delegation, agency, and voice were measured concurrently. Moreover, the indirect-effect coefficient and bootstrap confidence interval are absent from the present file; the 43.6% proportion mediated should not be emphasized until those verified estimates are reported. Any remaining direct association may reflect unmeasured confounding, measurement overlap, reverse directionality, or linguistic inheritance from model output.

The reported non-significant delegation-by-AI-literacy interaction does not support the proposed buffering role of AI literacy in this sample. AI literacy was positively correlated with self-perceived voice at the bivariate level ($r = .44$, $p < .001$) but was not a statistically significant unique predictor in the final regression model ($\beta = .148$, $p = .128$). Thus, the data do not justify the claim that AI literacy independently raised voice. A cautious pedagogical inference is that critical evaluation training may need to be combined with process guidance about when and how GenAI enters composing; this proposition requires longitudinal or experimental testing (UNESCO, 2023; Kohnke et al., 2023).

The corpus findings extend comparative text research (Amirjalili et al., 2024; Al Hosni, 2025) by linking individual differences in self-reported delegation to selected essay features within one learner sample. Delegation was associated with fewer self-mentions and hedges and more boosters. The strong correlation with self-mention ($r = -.76$) may be pedagogically informative at the aggregate level, but it is not evidence of authorship for an individual text and must not be used as an AI-detection criterion. Claims concerning sentence length and lexical diversity should be retained only after their inferential statistics are added to the Results. The broader assessment implication is that surface fluency and authorial presence are not interchangeable constructs.

The Pakistani setting sharpens the general finding. Delegation did not vary by prior medium of schooling or by level of study, which contradicts the assumption that ghost-writing is a coping strategy of the linguistically weakest students. Instead, it appears to be a time-and-risk strategy available to everyone and taken up broadly, consistent with the high general adoption reported by Rashid et al. (2025) and the strong trust-improvement associations documented by Anwer et al. (2025). The fourth qualitative theme adds something to that literature that the perception surveys could not reach: in an assessment culture that reads personal expression as unprofessional and as evidence of AI use, students suppress their voice deliberately. Institutional response to GenAI is thus itself contributing to the erosion it aims to prevent.

6. Pedagogical and Assessment Implications

Writing instruction should distinguish editorial assistance, dialogic support, and generative delegation rather than treating all GenAI use as equivalent. Process-focused assessment can require planning notes, prompt and revision logs, annotated drafts, brief oral justification of compositional decisions, or reflective disclosure of how model suggestions were accepted, rejected, and rewritten. Rubrics should assess ownership of claims, warranted stance, and revision reasoning alongside linguistic accuracy. These recommendations are provisional: the current study identifies associations in one regional convenience sample and does not test an intervention.

7. Conclusion

This study found that modes of GenAI use were differentially associated with authorial voice among 92 Pakistani university ESL learners. Editorial use was not significantly related to the measured voice outcomes, dialogic use showed positive bivariate associations, and generative delegation was associated with lower self-perceived and rubric-scored voice. Writer agency had the largest standardized coefficient in the final regression model, whereas AI literacy did not significantly moderate the delegation-voice association. Delegation was also associated with fewer self-mentions and hedges and more boosters in

submitted essays. Interviews suggested that participants often delegated the translation of ideas into prose, experienced a fluency-ownership trade-off, bypassed rough drafting, and managed voice in response to perceived institutional risk. These results support distinguishing assistance from delegation in pedagogy and assessment, but they do not show that GenAI use causes voice loss. Verified mediation and moderation coefficients, complete ethics approval details, reproducible corpus procedures, scale-validation evidence, and confidence intervals must be added before journal submission.

References

- Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21, Article 10. <https://doi.org/10.1186/s41239-024-00444-7>
- Al Hosni, J. K. (2025). Preserving authorial voice in academic texts in the age of generative AI: A thematic literature review. *Arab World English Journal*, 16(3), 244-258. <https://doi.org/10.24093/awej/vol16no3.14>
- Alghamdi, A. A. (2025). University students' perceptions of generative AI for critical thinking and creativity: The influence of self-efficacy and disciplinary differences. *Innovations in Education and Teaching International*, 1-15. <https://doi.org/10.1080/14703297.2025.2600476>
- Alghazo, R., Fatima, G., Malik, M., Abdelhamid, S. E., Jahanzaib, M., Nayab, D., & Raza, A. (2025). Exploring ChatGPT's role in higher education: Perspectives from Pakistani university students on academic integrity and ethical challenges. *Education Sciences*, 15(2), Article 158. <https://doi.org/10.3390/educsci15020158>
- Alkamel, M. A. A., & Alwagieh, N. A. S. (2024). Utilizing an adaptable artificial intelligence writing tool (ChatGPT) to enhance academic writing skills among Yemeni university EFL students. *Social Sciences & Humanities Open*, 10, Article 101095. <https://doi.org/10.1016/j.ssaho.2024.101095>
- Almassaad, A., Alajlan, H., & Alebaikan, R. (2024). Student perceptions of generative artificial intelligence: Investigating utilization, benefits, and challenges in higher education. *Systems*, 12(10), Article 385. <https://doi.org/10.3390/systems12100385>
- Alzubi, A. A. F., Nazim, M., & Alyami, N. (2025). Do AI-generative tools kill or nurture creativity in EFL teaching and learning? *Education and Information Technologies*, 30(11), 15147-15184. <https://doi.org/10.1007/s10639-025-13409-8>
- Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English literature. *Frontiers in Education*, 9, Article 1347421. <https://doi.org/10.3389/educ.2024.1347421>
- Anwer, B., Brohi, R., & Alvi, A. (2025). Trust in ChatGPT and perceived academic writing improvement: A TAM-based quantitative study in a Pakistani ESL context. *Futurity Education*, 5(4), 4-20.
- Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Information Age Publishing.
- 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. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qpo630a>
- Canagarajah, S. (2015). "Blessed in my own way": Pedagogical affordances for dialogical voice construction in multilingual student writing. *Journal of Second Language Writing*, 27, 122-139. <https://doi.org/10.1016/j.jslw.2014.09.001>
- Chen, Y., Zhang, X., & Liu, H. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research*, 23, Article 24. <https://doi.org/10.28945/5370>
- Cheng, A., Calhoun, A., & Reedy, G. (2025). Artificial intelligence-assisted academic writing: Recommendations for ethical use. *Advances in Simulation*, 10, Article 22.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Derakhshan, A., & Ghiasvand, F. (2024). Is ChatGPT an evil or an angel for second language education and research? A phenomenographic study of research-active EFL teachers' perceptions. *International Journal of Applied Linguistics*, 34(4), 1246-1264.
- Dong, L. (2025). Development and validation of a multidimensional ChatGPT feedback engagement scale in second language writing. *System*, 135, Article 103875.
- Elbow, P. (Ed.). (1994). *Landmark essays on voice and writing*. Hermagoras Press.
- Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365-387. <https://doi.org/10.2307/356600>
- Franjic, M., & Bozinovski, B. (2025). Exploring ChatGPT's impact on ESL writing proficiency: An experimental approach. *Teaching English with Technology*, 25(2), 53-81.

- Hayes, A. F. (2022). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (3rd ed.). Guilford Press.
- Hongxia, H., & Razali, A. B. (2025). Impact of ChatGPT on English academic writing ability and engagement of Chinese EFL undergraduates. *International Journal of Instruction*, 18(2), 323-346.
- 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. (2008). Disciplinary voices: Interactions in research writing. *English Text Construction*, 1(1), 5-22. <https://doi.org/10.1075/etc.1.1.03hyl>
- Ivanic, R. (1998). Writing and identity: The discursual construction of identity in academic writing. John Benjamins.
- Ivanic, R., & Camps, D. (2001). I am how I sound: Voice as self-representation in L2 writing. *Journal of Second Language Writing*, 10(1-2), 3-33. [https://doi.org/10.1016/S1060-3743\(01\)00034-0](https://doi.org/10.1016/S1060-3743(01)00034-0)
- Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kiziltas, Y. (2025). Integration of artificial intelligence (AI) into primary school students' writing skills: The impact of ChatGPT on creative writing and writing self-efficacy. *Journal of Educational Computing Research*, 63(7-8), 1717-1747. <https://doi.org/10.1177/07356331251365187>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>
- Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Article 1121). ACM. <https://doi.org/10.1145/3706598.3713778>
- Levine, S., Beck, S. W., Mah, C., Phalen, L., & Pittman, J. (2025). How do students use ChatGPT as a writing support? *Journal of Adolescent & Adult Literacy*, 68(5), 445-457. <https://doi.org/10.1002/jaal.1373>
- 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](https://doi.org/10.1016/S1060-3743(00)00036-9)
- Matsuda, P. K., & Tardy, C. M. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes*, 26(2), 235-249. <https://doi.org/10.1016/j.esp.2006.10.001>
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2), Article 07. <https://doi.org/10.53761/1.20.02.07>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879-891. <https://doi.org/10.3758/BRM.40.3.879>
- Prior, P. (2001). Voices in text, mind, and society: Sociohistoric accounts of discourse acquisition and use. *Journal of Second Language Writing*, 10(1-2), 55-81. [https://doi.org/10.1016/S1060-3743\(00\)00037-0](https://doi.org/10.1016/S1060-3743(00)00037-0)
- Rashid, S., Malik, S., Abbas, F., & Ahmad, M. (2025). Pakistani students' perceptions about knowledge, use and impact of artificial intelligence (AI) on academic writing: A case study. *Journal of Computers in Education*, 12, 1053-1078. <https://doi.org/10.1007/s40692-024-00338-7>
- Ryan, H., Abramov, D., Acker, S., & Elkins, S. (2025). Can AI be a co-author? How generative AI challenges the boundaries of authorship in a general education writing class. *Journal of Teaching and Learning*, 19(1), 40-57.
- Seran, C. E., Tan, M. J. T., Karim, H. A., & AlDahoul, N. (2025). A conceptual exploration of generative AI-induced cognitive dissonance and its emergence in university-level academic writing. *Frontiers in Artificial Intelligence*, 8, Article 1573368. <https://doi.org/10.3389/frai.2025.1573368>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, Article 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Stapleton, P. (2002). Critiquing voice as a viable pedagogical tool in L2 writing: Returning the spotlight to ideas. *Journal of Second Language Writing*, 11(3), 177-190. [https://doi.org/10.1016/S1060-3743\(02\)00070-X](https://doi.org/10.1016/S1060-3743(02)00070-X)
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, Article 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Tardy, C. M. (2012). Current conceptions of voice. In K. Hyland & C. S. Guinda (Eds.), *Stance and voice in written academic genres* (pp. 34-48). Palgrave Macmillan.
- UNESCO. (2023). *Guidance for generative AI in education and research* (F. Miao & W. Holmes, Authors). United Nations Educational, Scientific and Cultural Organization.

- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for learners of English as a second or foreign language. *Journal of Second Language Writing*, 62, Article 101071. <https://doi.org/10.1016/j.jslw.2023.101071>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943-13967. <https://doi.org/10.1007/s10639-023-11742-4>
- 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., & Llosa, L. (2008). Voice in high-stakes L1 academic writing assessment: Implications for L2 writing instruction. *Assessing Writing*, 13(3), 153-170. <https://doi.org/10.1016/j.asw.2008.10.003>