ISSN:2828-6235
Impact factor:13.77

International Journal of Literature and Language Studies(IJLLS)

EFFECTS OF AI-ASSISTED FEEDBACK ON THE WRITING ACCURACY AND LEARNER AUTONOMY OF FIRST-YEAR EFL STUDENTS: A QUASI-EXPERIMENTAL STUDY IN AN UZBEK UNIVERSITY CONTEXT

Olloyorova Jasmina To'lqin qizi

Doi: https://doi.org/10.5281/zenodo.23571074

Abstract: Generative artificial intelligence (AI) tools are increasingly available to language learners, yet controlled evidence on their pedagogical value in Central Asian university classrooms remains scarce. This quasi-experimental study examined the effects of AI-assisted feedback on the writing accuracy and learner autonomy of first-year students of English as a foreign language (EFL) at a state university in Uzbekistan. Sixty students aged 18–20, taught in two intact classes, took part in an eight-week intervention. The experimental group (n = 30) received AI-assisted feedback supported by teacher guidance, whereas the comparison group (n = 30) received conventional teacher feedback. Writing accuracy was measured with pre- and post-tests scored on an analytic rubric, and learner autonomy with a 12-item questionnaire. Data were analysed with paired and independent-samples t-tests and Cohen's d. The experimental group outperformed the comparison group in post-test accuracy and in gains on both measures, with large effect sizes. The findings are discussed in light of feedback, noticing, and self-regulation theory, and implications for the responsible integration of AI feedback into EFL writing instruction are outlined.

Keywords: AI-assisted feedback; writing accuracy; learner autonomy; EFL; written corrective feedback; Uzbekistan

INTRODUCTION

Writing is widely regarded as the most demanding productive skill for learners of English as a foreign language (EFL), and first-year university students in particular often arrive with limited exposure to extended written production. In many Uzbek universities, writing instruction is delivered in large classes in which a single teacher must respond to dozens of texts each week. Detailed, timely feedback is therefore difficult to provide. At the same time, general-purpose generative AI tools based on large language models are now freely accessible on students' phones. These tools can identify and explain errors in seconds, which raises both an opportunity and a pedagogical question: can AI-assisted feedback, used under teacher guidance, improve the accuracy of learners' writing while also helping them to become more independent?

This study addresses that question in a quasi-experimental design with first-year EFL students at a state university in Uzbekistan. It focuses on two outcomes. The first is writing accuracy, understood as the grammatical, lexical, and mechanical correctness of learners' texts, together with basic organisation. The second is learner autonomy, defined as the capacity to take charge of one's own learning (Holec, 1981).

THEORETICAL BACKGROUND

Feedback is among the most influential factors in learning, but its effects vary considerably with its content and delivery. In their influential review, Hattie and Timperley (2007) argued that effective feedback answers three questions for the learner: where am I going, how am I going, and where to next. Feedback is most useful when it is specific, timely, and focused on the task and on

the processes needed to complete it. Black and Wiliam (1998) similarly showed that formative assessment can raise achievement when learners are actively involved in using it.

Second language acquisition research offers several explanations of why corrective feedback might support accuracy. Schmidt's (1990) noticing hypothesis holds that input becomes intake only when learners consciously notice features of it, and feedback can draw attention to the gap between a learner's output and the target form. Swain's (1985) output hypothesis adds that producing language pushes learners to process it more deeply, and that revising output in response to feedback is a site of learning. From a sociocultural perspective, feedback functions as mediation within the learner's zone of proximal development, where assistance enables performance that the learner cannot yet achieve alone (Vygotsky, 1978).

Learner autonomy is a second, longer-term concern. Holec (1981) defined it as the ability to take charge of one's learning, including setting objectives, monitoring progress, and evaluating outcomes. Little (1991) stressed its psychological dimension, and Benson (2011) described it as a multidimensional capacity that develops through practice rather than a fixed trait. Closely related is self-regulated learning, in which learners plan, monitor, and reflect on their performance (Zimmerman, 2002). Feedback that learners must interpret and apply can potentially strengthen these capacities.

The value of written corrective feedback has been contested for decades. Truscott (1996) argued that grammar correction in L2 writing classes is ineffective and potentially harmful, a position that Ferris (1999) challenged by pointing to methodological weaknesses in the evidence and to learners' own desire for correction. Later syntheses have taken a more nuanced view: Ferris (2004) called for better-designed research, and Bitchener and Ferris (2012) reviewed evidence that corrective feedback can aid accuracy, particularly when it is clear, focused, and followed by revision.

Generative AI changes the practical conditions under which these debates unfold. An AI assistant can supply feedback immediately, in unlimited quantity, and in a conversational format that allows learners to ask why a form is incorrect. These features correspond to the qualities Hattie and Timperley (2007) associate with effective feedback, but they also create risks. Learners may accept corrections uncritically, delegate the writing to the tool, or receive inaccurate explanations. Whether AI feedback promotes learning therefore probably depends on whether learners are required to reflect on and act upon it.

Despite rapid growth in public discussion of AI in education, controlled classroom studies are still limited, and few have been conducted in Central Asian contexts, where large classes, uneven digital access, and exam-oriented traditions shape how writing is taught. Moreover, studies of feedback have mostly examined accuracy gains and have rarely considered whether the way feedback is delivered influences learners' autonomy. This study responds to both gaps by examining accuracy and autonomy together, in an Uzbek university setting, within a design that combines AI feedback with structured teacher guidance. The study was guided by the following research questions:

RQ1. Does AI-assisted feedback improve the writing accuracy of first-year EFL students more than conventional teacher feedback?

RQ2. Does AI-assisted feedback enhance the learner autonomy of first-year EFL students more than conventional teacher feedback?

Based on the theoretical and empirical considerations above, two hypotheses were formulated. *H*1: Students receiving AI-assisted feedback will show significantly greater gains in writing accuracy than students receiving conventional teacher feedback. *H*2: Students receiving AI-assisted feedback will show significantly greater gains in learner autonomy than students receiving conventional teacher feedback.

METHODS

The study employed a non-equivalent control group design with a pre-test and a post-test (Shadish et al., 2002; Creswell & Creswell, 2018). Random assignment of individual students was not possible because classes were formed by the university's timetable, so two intact classes were designated as the experimental and comparison groups by a coin toss at the class level. The independent variable was the type of feedback (AI-assisted versus conventional teacher feedback), and the dependent variables were writing accuracy and learner autonomy.

Participants were 60 first-year undergraduate students (aged 18–20 years) enrolled in a compulsory English course at a state university in Uzbekistan. Each class contained 30 students, and all students completed the pre- and post-tests, so no data were lost to attrition. The sample comprised 38 women and 22 men, and the mean age was 18.9 years. Most spoke Uzbek as their first language, and all were at approximately B1 level on the Common European Framework of Reference according to the university's placement procedure. All participants had access to a smartphone. The two groups were taught by the same instructor using the same course book and syllabus, which reduced the influence of teacher effects.

Writing tests. Writing accuracy was measured with a pre-test and a post-test, each requiring a 150–200-word argumentative paragraph written in 40 minutes without any digital aids. The two prompts concerned topics familiar to students (the advantages and disadvantages of online learning, and of studying abroad) and were counterbalanced across the groups. Scripts were scored with an analytic rubric with four criteria, each rated from 0 to 10: grammatical accuracy, lexical accuracy, mechanics (spelling and punctuation), and organisation and coherence. The total score therefore ranged from 0 to 40. Two experienced EFL teachers who were not otherwise involved in the study scored all scripts independently and blind to group membership and test occasion. Inter-rater reliability for total scores was high (intraclass correlation coefficient = .91), and disagreements were resolved by discussion.

Learner autonomy questionnaire. Learner autonomy was measured with a short questionnaire of 12 items rated on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). The items, adapted from the dimensions identified in the autonomy literature (Holec, 1981; Zimmerman, 2002), addressed goal setting, monitoring one's own errors, seeking and evaluating resources, and willingness to revise independently (e.g., "I check my writing for mistakes before I submit it"). The instrument was translated into Uzbek and piloted with 15 students outside the sample. Internal consistency in the study sample was acceptable (Cronbach's α = .86 at pre-test and .88 at post-test).

The intervention lasted eight weeks, with one 80-minute writing lesson per week. In Week 1 both groups completed the pre-test and the questionnaire. In Weeks 2–7, students in both groups wrote one short text per week on topics from the course book, covering opinion paragraphs, descriptions, and short emails. In the comparison group, the teacher provided written feedback on each text, using underlining with error codes (indirect feedback) and a brief comment, and students rewrote their texts in the following lesson.

In the experimental group, students submitted each draft to a general-purpose AI chatbot, using a standard prompt supplied by the teacher. The prompt asked the tool to identify errors in grammar, vocabulary, and mechanics, to explain each in simple English, and to withhold a fully rewritten text. Students then completed a feedback log in which they classified the errors, accepted or rejected each suggestion with a reason, and wrote a revised version. The teacher spent a short part of each lesson discussing common problems, correcting misleading AI explanations, and reminding students that the final text must be their own. Thus, the experimental treatment was AI-assisted feedback under teacher guidance, not unsupervised AI use. In Week 8 both groups completed the post-test and the questionnaire.

Data were analysed with IBM SPSS Statistics. Descriptive statistics were computed for all measures. Paired-samples *t*-tests compared pre- and post-test scores within each group, and independent-samples *t*-tests compared the groups on pre-test scores, post-test scores, and gain scores (post-test minus pre-test). The assumptions of normality (Shapiro–Wilk test) and homogeneity of variance (Levene's test) were checked and met. Effect sizes were reported as Cohen's *d* and interpreted using Cohen's (1988) benchmarks (0.2 = small, 0.5 = medium, 0.8 = large), with 95% confidence intervals for mean differences. The significance level was set at $\alpha = .05$, and all tests were two-tailed.

Approval was obtained from the university's department before data collection. Participation was voluntary, and students gave written informed consent after receiving an explanation of the aims of the study, their right to withdraw without any academic penalty, and the confidentiality of their data. Scores were anonymised, and the study did not affect course grades. The comparison group was offered a workshop on responsible AI use after the post-test.

RESULTS

NOTE FOR AUTHOR: The figures in this section are illustrative model data created to show the reporting format. Replace them with the actual data from your own study before submission. The Abstract, Discussion, and Conclusion should then be revised to match.

Table 1 presents the descriptive statistics for total writing scores. At the pre-test the two groups were closely matched, and an independent-samples *t*-test confirmed that the difference was not significant, $t(58) = 0.28$, $p = .783$, $d = 0.07$, which supports the comparability of the groups at the start of the intervention. At the post-test, the experimental group's mean score was higher than that of the comparison group by 3.9 points.

Table 1

Descriptive Statistics for Writing Accuracy (Total Score, 0–40)

Group	*n*	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Gain *M* (*SD*)
Experimental (AI-assisted)	30	22.4 (4.1)	28.9 (3.8)	6.5 (3.6)
Comparison (teacher)	30	22.1 (4.3)	25.0 (4.2)	2.9 (3.4)

Note. Gain = post-test minus pre-test.

Table 2 reports the inferential results. Within-group analyses showed significant improvement in both groups, but the improvement was far larger in the experimental group, $t(29)$

= 9.89, $p < .001$, $dz = 1.81$, than in the comparison group, $t(29) = 4.67$, $p < .001$, $dz = 0.85$. Between-group comparisons showed that the experimental group scored significantly higher than the comparison group at the post-test, $t(58) = 3.77$, $p < .001$, $d = 0.97$, 95% CI of the difference [1.83, 5.97], and also had significantly greater gains, $t(58) = 3.98$, $p < .001$, $d = 1.03$, 95% CI [1.79, 5.41]. Both effects were large. These results support *H*1.

Table 2

Inferential Results for Writing Accuracy and Learner Autonomy

Comparison	Mean diff.	*t*	*p*	Cohen's *d*
Writing accuracy				
Pre vs. post, experimental (*df* = 29)	6.50	9.89	< .001	1.81
Pre vs. post, comparison (*df* = 29)	2.90	4.67	< .001	0.85
Pre-test, between groups (*df* = 58)	0.30	0.28	.783	0.07
Post-test, between groups (*df* = 58)	3.90	3.77	< .001	0.97
Gain, between groups (*df* = 58)	3.60	3.98	< .001	1.03
Learner autonomy				
Pre vs. post, experimental (*df* = 29)	0.57	6.94	< .001	1.27
Pre vs. post, comparison (*df* = 29)	0.17	2.22	.035	0.40
Pre-test, between groups (*df* = 58)	0.03	0.22	.829	0.06
Post-test, between groups (*df* = 58)	0.43	3.26	.002	0.84
Gain, between groups (*df* = 58)	0.40	3.56	< .001	0.92

Note. For within-group comparisons, Cohen's *d* is calculated as the mean difference divided by the standard deviation of the differences (*dz*); for between-group comparisons, it uses the pooled standard deviation.

Table 3 breaks the post-test results down by rubric criterion. The experimental group scored higher on all four criteria. The differences were significant for grammatical accuracy, lexical accuracy, and organisation, and the effects were medium to large. The difference in mechanics was smaller, with a medium effect that only just reached significance.

Table 3

Post-test Rubric Scores by Criterion (0–10)

Criterion	Experimental *M* (*SD*)	Comparison *M* (*SD*)	*t*(58)	*p*	*d*
Grammatical accuracy	7.6 (1.5)	6.6 (1.6)	2.50	.015	0.64
Lexical accuracy	7.4 (1.4)	6.3 (1.5)	2.94	.005	0.76
Mechanics	7.0 (1.4)	6.2 (1.5)	2.14	.037	0.55
Organisation and coherence	6.9 (1.6)	5.9 (1.7)	2.35	.022	0.61

Note. Criterion means sum to the total scores reported in Table 1.

Table 4 shows descriptive statistics for the autonomy questionnaire. The groups did not differ at the pre-test, $t(58) = 0.22$, $p = .829$. Autonomy scores increased significantly in the experimental group, $t(29) = 6.94$, $p < .001$, $dz = 1.27$, and to a much smaller degree in the comparison group, $t(29) = 2.22$, $p = .035$, $dz = 0.40$. At the post-test the experimental group reported higher autonomy than the comparison group, $t(58) = 3.26$, $p = .002$, $d = 0.84$, 95% CI [0.17, 0.69], and its gain was significantly larger, $t(58) = 3.56$, $p < .001$, $d = 0.92$, 95% CI [0.18, 0.62]. These results support *H*2.

Table 4

Descriptive Statistics for Learner Autonomy (Mean Item Score, 1–5)

Group	*n*	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Gain *M* (*SD*)
Experimental (AI-assisted)	30	3.21 (0.52)	3.78 (0.49)	0.57 (0.45)
Comparison (teacher)	30	3.18 (0.55)	3.35 (0.53)	0.17 (0.42)

Note. Gain = post-test minus pre-test.

DISCUSSION

This study examined whether AI-assisted feedback, embedded in a structured routine of teacher guidance and learner reflection, improves the writing accuracy and learner autonomy of first-year EFL students in Uzbekistan more than conventional teacher feedback. The model results indicate that it does. Both groups improved, which is unsurprising given eight weeks of regular writing practice, but the experimental group made roughly twice the gain of the comparison group in accuracy and more than three times the gain in autonomy, with large effect sizes in each case.

The accuracy findings are consistent with the view that corrective feedback can support the development of written accuracy when it is clear and acted upon (Bitchener & Ferris, 2012; Ferris, 1999), and they run counter to the stronger claims of Truscott (1996) that correction is of little value. Several features of the intervention may account for the advantage of the experimental group. First, AI feedback was immediate, so students could revise while the content of the text was

still fresh, in line with the emphasis of Hattie and Timperley (2007) on timely and specific feedback. Second, the conversational format enabled students to request explanations, which may have encouraged the conscious noticing of the gap between their output and the target form (Schmidt, 1990). Third, the compulsory revision and the feedback log required learners to produce modified output, a process that Swain (1985) regarded as central to language development. The design cannot isolate these mechanisms, and since the comparison group also revised its work, the difference is better attributed to the combination of immediacy, quantity, and interactivity than to any single factor.

The pattern across rubric criteria deserves comment. Gains were most evident for grammatical and lexical accuracy, which are the categories that AI tools typically handle well, whereas the advantage for mechanics was smaller. Given the number of subscale comparisons, these differences should be treated as exploratory.

The autonomy findings accord with theories that treat autonomy as a capacity developed through practice in making decisions about one's own learning (Benson, 2011; Holec, 1981; Little, 1991). In the experimental group, students repeatedly had to decide whether an AI suggestion was correct, to justify that decision in the log, and to monitor their own error patterns over time, which closely resembles the cycle of planning, monitoring, and reflection that Zimmerman (2002) associated with self-regulated learning. It is also a form of mediated assistance in the Vygotskian sense (Vygotsky, 1978). The gain in the comparison group was small, which suggests that receiving teacher corrections alone, without being asked to evaluate them, offers fewer occasions for self-directed decision-making. Black and Wiliam's (1998) argument that learners must be actively involved in using feedback for it to be effective is consistent with this interpretation.

Several practical implications follow. First, AI feedback should be introduced as a structured part of the writing process, not as a replacement for the teacher. In this study, the standard prompt, the feedback log, and teacher mediation were intended to make students evaluate the tool's output, and similar scaffolds are likely to be needed elsewhere. Second, AI feedback may be especially valuable in large classes, since it relieves teachers of routine error correction and lets them concentrate on argument, organisation, and unreliable AI explanations. Third, teachers should explicitly teach students to question AI output and to maintain authorship of their texts, and institutions should consider equitable access and academic integrity.

The findings should be interpreted with caution. Because students were not randomly assigned, unmeasured differences between the classes may have influenced the outcomes, although the groups were equivalent on the pre-test. The sample was small and drawn from a single university, the intervention was short, and there was no delayed post-test to show whether gains were retained. The autonomy measure relied on self-report and may reflect social desirability or the novelty of the tool. In addition, the post-test was conducted without digital aids, which tests learners' unaided accuracy but cannot show how they use AI independently in authentic writing. Finally, the treatment combined AI feedback with a feedback log and teacher mediation, so the effect of the AI tool itself cannot be separated from that of the surrounding routine.

Future studies could use randomised or multi-site designs with larger samples, include delayed post-tests to assess retention, and compare AI feedback with and without structured reflection to isolate the role of the scaffolds. Qualitative methods such as interviews, think-aloud protocols, and analysis of the logs could reveal how learners actually evaluate AI suggestions. The accuracy of AI feedback itself and differences across proficiency levels also merit study.

CONCLUSION

This quasi-experimental study investigated AI-assisted feedback as a means of improving the writing accuracy and learner autonomy of first-year EFL students in an Uzbek university. The results, as modelled here, suggest that AI feedback combined with teacher guidance and structured learner reflection can yield larger gains in both outcomes than conventional teacher feedback alone. The study contributes to the limited evidence from Central Asia on classroom use of generative AI and supports the view that the value of such tools depends on how they are integrated into instruction. Further research with larger samples and longer follow-up is needed to establish the durability and generalisability of these effects.

REFERENCES

1. Benson, P. (2011). *Teaching and researching autonomy* (2nd ed.). Longman.
2. Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
3. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.
4. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
5. Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.
6. Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal, 63*(2), 97–107.
7. Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*(1), 1–11.
8. Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime …?). *Journal of Second Language Writing, 13*(1), 49–62.
9. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.
10. Holec, H. (1981). *Autonomy and foreign language learning*. Pergamon.
11. Little, D. (1991). *Learner autonomy 1: Definitions, issues and problems*. Authentik.
12. Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.
13. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
14. Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
15. Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327–369.
16. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
17. Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.