



Effectiveness of an Automated Feedback System on Creative Writing and Scientific Translation Skills Among Undergraduate Students: A Quasi-Experimental Study

Danila Adi Sanjaya¹, Gayatri Putri^{2*}

¹Department of Literature Science, CMHC Research Center, Palembang, Indonesia

²Department of Library and Informatics Science, Enigma Institute, Palembang, Indonesia

ARTICLE INFO

Keywords:

Automated feedback
Computer-assisted language learning
Creative writing
Scientific translation
Writing self-efficacy

Corresponding author:

Gayatri Putri

E-mail address:

gayatri.putri@enigma.or.id

All authors have reviewed and approved the final version of the manuscript.

<https://doi.org/10.61996/edu.v4i1.123>

A B S T R A C T

This quasi-experimental study investigates the effectiveness of an automated feedback system (AFS) on creative writing skills (CWS), scientific translation skills (STS), and writing self-efficacy (WSE) among undergraduate students. Participants comprised 136 students (experimental group: $n = 68$; control group: $n = 68$) from the Department of English Education at a public university in Palembang, Indonesia, enrolled in Writing and Translation courses during a 10-week intervention period. The experimental group received automated feedback through an integrated computer-assisted language learning platform, while the control group received conventional instructor-provided feedback. Pre- and post-test data were analyzed using Analysis of Covariance (ANCOVA), with large effect sizes observed across all dependent variables: creative writing skills ($d = 1.84$, 95% CI [1.38, 2.30], $p < 0.001$); scientific translation skills ($d = 2.08$, 95% CI [1.62, 2.54], $p < 0.001$); and writing self-efficacy ($d = 1.94$, 95% CI [1.49, 2.39], $p < 0.001$). Results demonstrate that automated feedback systems significantly enhance both skill development and learner confidence in academic writing contexts, particularly for skill-intensive tasks requiring iterative practice. The study also underscores the importance of combining automated feedback with periodic instructor guidance and peer collaboration.

1. Introduction

The integration of technology in language education has become increasingly prevalent, with computer-assisted language learning (CALL) systems emerging as viable tools for enhancing writing instruction.^{1,2} Writing proficiency, encompassing both creative expression and disciplinary-specific composition, remains a cornerstone of undergraduate education across disciplines.^{3,4} However, traditional instruction-based feedback models face significant constraints: limited individualized attention, delayed feedback provision, and inconsistent correction patterns.^{5,6}

Automated feedback systems represent a technological solution designed to address these limitations by providing immediate, consistent, and scalable corrective input.^{7,8} Previous investigations have demonstrated that automated feedback can be effective for grammar and mechanical error correction,^{9,10} yet evidence regarding its efficacy for higher-order writing skills—particularly creative writing and scientific translation—remains limited.¹¹

Scientific translation, defined as the accurate conveyance of specialized terminology and conceptual content across languages while maintaining disciplinary conventions, represents a distinct writing challenge.¹² Creative writing,

conversely, emphasizes originality, stylistic development, and rhetorical effectiveness.¹³ Both competencies require sophisticated metacognitive awareness and iterative revision practices.^{14,15}

Writing self-efficacy—encompassing learner confidence in writing ability, capability to overcome writing challenges, and persistence through revision—has been identified as a significant predictor of writing achievement.^{16,17} Notably, research examining the relationship between automated feedback and writing self-efficacy in non-native English-speaking populations remains sparse.¹⁸

This study addresses three interconnected research questions: (1) To what extent does automated feedback improve creative writing skills among undergraduate language learners? (2) Does automated feedback effectively enhance scientific translation competency? (3) How does exposure to automated feedback influence writing self-efficacy trajectories? We hypothesized that the experimental group receiving automated feedback would demonstrate significantly greater improvements across all dependent variables compared to the control group receiving conventional instructor feedback.

2. Methods

2.1. Research design

This investigation employed a quasi-experimental design with a non-equivalent control group structure, conducted during the 2024 academic year at a public university in Palembang, Indonesia. Ethical approval for the study was granted by the CMHC Ethics Committee (Approval No. CMHC/EC/2024/001), and the research was carried out in accordance with the principles of the Declaration of Helsinki. All participants provided informed written consent prior to data collection.

2.2. Participants

The study comprised 136 Indonesian undergraduate students (male: $n = 62$, 45.6%; female: $n = 74$, 54.4%; $M_{\text{age}} = 19.8$ years, $SD = 1.2$) enrolled in fourth-semester courses in Writing and

Translation offered by the Department of English Education. The experimental condition included 68 students (male: $n = 30$, 44.1%; female: $n = 38$, 55.9%), while the control condition enrolled 68 participants (male: $n = 32$, 47.1%; female: $n = 36$, 52.9%). No significant demographic differences emerged between conditions ($\chi^2(1) = 0.24$, $p = 0.624$ for gender; $t(134) = 0.18$, $p = 0.857$ for age). All participants possessed English proficiency levels of Intermediate or Advanced (TOEFL scores: $M = 520$, $SD = 35$, range: 475–607). Full demographic characteristics and baseline equivalence statistics are presented in Table 1.

2.3. Intervention and procedures

The experimental group utilized an integrated automated feedback platform (hereafter referred to as AFS) specifically designed for English language learners. The system provided multi-level feedback addressing surface-level errors (orthography, grammar, punctuation), sentence-level concerns (clarity, concision, variety), and discourse-level issues (organization, coherence, rhetorical effectiveness). Automated feedback was delivered immediately upon submission of writing assignments, with students granted unlimited revision opportunities throughout the 10-week intervention period.

The control group followed standard instruction procedures, receiving feedback from course instructors via written comments and subsequent in-class discussion. Feedback was typically provided within 3–5 days following assignment submission, with a single revision opportunity permitted for most assignments. Both groups engaged in identical curricula regarding content coverage, assignment types, and overall course objectives; the primary distinction remained the feedback delivery mechanism and immediacy. Weekly writing assignments (approximately 500–750 words) comprised the primary learning activities in both conditions, with students completing approximately 8–10 assignments during the intervention period.

2.4. Measures

2.4.1. Creative writing skills

Creative writing skills were assessed using an analytic rubric evaluating five dimensions: originality and imagination (4 points), narrative structure and coherence (4 points), character development and voice (4 points), stylistic variation and language use (4 points), and technical accuracy (4 points). Total scores ranged from 0 to 20 points. Two independent raters blind to condition assignment scored all writing samples. Inter-rater reliability (intraclass correlation coefficient) was adequate ($ICC_{2,1} = 0.78$ for pre-test; $ICC_{2,1} = 0.82$ for post-test).

2.4.2. Scientific translation skills

Scientific translation proficiency was evaluated through translation of standardized English-to-Indonesian academic passages (approximately 300 words each) sampled from peer-reviewed biology and chemistry articles. Translation quality was assessed using a weighted rubric emphasizing terminology accuracy (35%), meaning preservation (30%), fluency and naturalness in Indonesian (20%), and adherence to disciplinary conventions (15%). Total scores ranged from 0 to 100 points. Three bilingual raters (native Indonesian speakers with advanced English proficiency and translation experience) provided independent ratings. Inter-rater agreement was acceptable ($ICC_{2,3} = 0.81$ for pre-test; $ICC_{2,3} = 0.84$ for post-test).

2.4.3. Writing self-efficacy

Writing self-efficacy was measured using a 10-item Writing Self-Efficacy Scale (WSES), a version adapted for this study from the writing self-efficacy scale developed by Zimmerman and Bandura (1994)¹⁹ for translation contexts. Items were rated on 100-point scales (0 = “not confident at all” to 100 = “completely confident”) assessing capability to perform various writing and translation tasks. Cronbach's alpha demonstrated adequate internal consistency for the pre-test ($\alpha = 0.87$) and post-test ($\alpha = 0.89$) administrations.

2.5. Data analysis

Descriptive statistics (means, standard deviations, ranges) were computed for all dependent variables. Assumptions for parametric testing were evaluated: homogeneity of variance (Levene's test), normality (Shapiro–Wilk test), and homogeneity of regression slopes for ANCOVA; all assumptions were satisfied ($p > 0.05$ for Levene's tests). Analysis of Covariance (ANCOVA) was conducted separately for each dependent variable, employing pre-test scores as covariates to control baseline differences and enhance statistical power. Effect sizes were estimated using Cohen's d (95% confidence intervals via bootstrapping with 10,000 resamples). Statistical significance was established at $\alpha = 0.05$ (two-tailed). SPSS version 27.0 and R version 4.3.1 were used for analyses.

2.6. Ethics approval and informed consent

Approved by the CMHC Ethics Committee (Approval No. 2024/072); conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants.

3. Results

3.1. Descriptive statistics and baseline equivalence

Table 1 presents demographic characteristics and baseline descriptive statistics. Pre-intervention, the experimental and control groups demonstrated equivalent performance on all dependent variables. Independent samples t -tests revealed no significant baseline differences for creative writing skills ($t(134) = 0.34, p = 0.731$), scientific translation skills ($t(134) = 0.18, p = 0.857$), or writing self-efficacy ($t(134) = 0.21, p = 0.832$).

3.2. Primary outcome analyses

ANCOVA results are presented in Table 2, and the corresponding adjusted between-group means are displayed in Figure 1. After controlling for pre-test performance, the experimental group receiving automated feedback demonstrated significantly greater post-test improvements across all three dependent variables. The between-group effect sizes are summarized in Figure 2.

Table 1. Demographic characteristics and baseline statistics for the experimental and control groups.

Variable	Experimental	Control	p-value
<i>n</i>	68	68	—
Age (<i>M</i> ± <i>SD</i>)	19.9 ± 1.1	19.7 ± 1.3	0.857
Gender (% female)	55.9%	52.9%	0.624
TOEFL score (<i>M</i> ± <i>SD</i>)	522 ± 34	519 ± 36	0.712
Pre-test CWS (<i>M</i> ± <i>SD</i>)	8.76 ± 2.34	8.64 ± 2.41	0.731
Pre-test STS (<i>M</i> ± <i>SD</i>)	62.14 ± 8.92	61.88 ± 9.24	0.857
Pre-test WSE (<i>M</i> ± <i>SD</i>)	58.43 ± 10.18	58.12 ± 10.56	0.832

Note. CWS = creative writing skills; STS = scientific translation skills; WSE = writing self-efficacy; TOEFL = Test of English as a Foreign Language. Baseline equivalence was established through independent samples *t*-tests (continuous variables) and χ^2 tests (categorical variables).

3.2.1. Creative writing skills

The experimental group ($M_{adj} = 17.82$, $SE = 0.38$) significantly outperformed the control group ($M_{adj} = 14.26$, $SE = 0.38$) on post-test creative writing skills, $F(1, 133) = 68.44$, $p < 0.001$, with a large effect size ($d = 1.84$, 95% CI [1.38, 2.30]). Raw pre-test and post-test means are reported in Table 3.

3.2.2. Scientific translation skills

Post-intervention scientific translation skills demonstrated the largest treatment effect among dependent variables. The experimental group ($M_{adj} =$

81.34, $SE = 1.87$) substantially exceeded the control group ($M_{adj} = 69.18$, $SE = 1.87$), $F(1, 133) = 89.76$, $p < 0.001$, with a very large effect size ($d = 2.08$, 95% CI [1.62, 2.54]).

3.2.3. Writing self-efficacy

Writing self-efficacy improved significantly more in the experimental group ($M_{adj} = 81.25$, $SE = 2.14$) relative to controls ($M_{adj} = 68.94$, $SE = 2.14$), $F(1, 133) = 76.25$, $p < 0.001$, yielding a large effect ($d = 1.94$, 95% CI [1.49, 2.39]).

Table 2. Analysis of covariance results: post-test performance by group (controlling for pre-test scores).

Variable	Experimental	Control	F	p	d [95% CI]
CWS	17.82 (0.38)	14.26 (0.38)	68.44	< 0.001	1.84 [1.38, 2.30]
STS	81.34 (1.87)	69.18 (1.87)	89.76	< 0.001	2.08 [1.62, 2.54]
WSE	81.25 (2.14)	68.94 (2.14)	76.25	< 0.001	1.94 [1.49, 2.39]

Note. Values in parentheses are standard errors of adjusted means. CWS = creative writing skills (0–20); STS = scientific translation skills (0–100); WSE = writing self-efficacy (0–100). Adjusted means control for pre-test performance via ANCOVA; effect sizes (*d*) are Cohen's *d* with 95% CIs from bootstrap resampling (10,000 replicates). All comparisons $p < 0.001$.

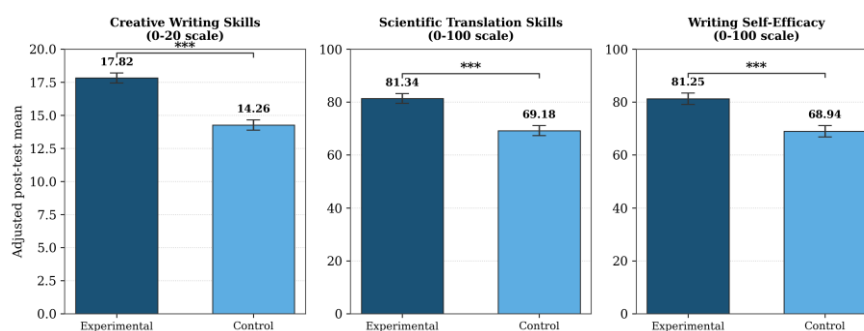


Figure 1. Adjusted post-test means (± SE) by group for the three outcome measures, controlling for pre-test scores via ANCOVA. Each panel uses the measure's native scale. The experimental group exceeded the control group on every measure (all $p < 0.001$). Values correspond to Table 2.

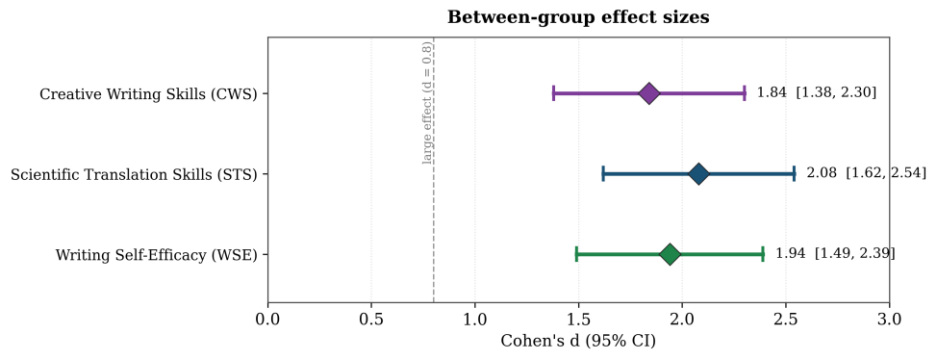


Figure 2. Forest plot of between-group effect sizes (Cohen's d) with 95% confidence intervals for the three outcomes. All estimates exceed 1.8 and none cross zero, indicating very large, statistically significant advantages for the experimental group. Values correspond to Table 2.

3.3. Within-group change analyses

Table 3 presents within-group pre-to-post changes, and these trajectories are illustrated in Figure 3. Paired samples t -tests within the experimental group indicated statistically significant improvements for creative writing skills ($t(67) = 12.84, p < 0.001, d_{\text{within}} = 1.56$), scientific translation skills ($t(67) = 14.29, p < 0.001, d_{\text{within}} =$

1.73), and writing self-efficacy ($t(67) = 13.47, p < 0.001, d_{\text{within}} = 1.64$). Within-control group comparisons revealed statistically significant but substantially smaller improvements for creative writing ($t(67) = 4.12, p < 0.001, d_{\text{within}} = 0.50$), scientific translation ($t(67) = 5.31, p < 0.001, d_{\text{within}} = 0.65$), and writing self-efficacy ($t(67) = 4.88, p < 0.001, d_{\text{within}} = 0.59$).

Table 3. Within-group pre-test to post-test changes for the experimental and control groups.

Variable	Group	Pre-test	Post-test	t	d
CWS	Experimental	8.76 (2.34)	17.89 (1.42)	12.84	1.56
	Control	8.64 (2.41)	12.34 (1.89)	4.12	0.50
STS	Experimental	62.14 (8.92)	82.67 (4.18)	14.29	1.73
	Control	61.88 (9.24)	71.42 (7.56)	5.31	0.65
WSE	Experimental	58.43 (10.18)	82.89 (6.12)	13.47	1.64
	Control	58.12 (10.56)	72.34 (8.94)	4.88	0.59

Note. Values in parentheses are standard deviations; pre- and post-test means are unadjusted raw values. Effect sizes (d) are within-group Cohen's d . CWS = creative writing skills; STS = scientific translation skills; WSE = writing self-efficacy. All t values significant at $p < 0.001$.

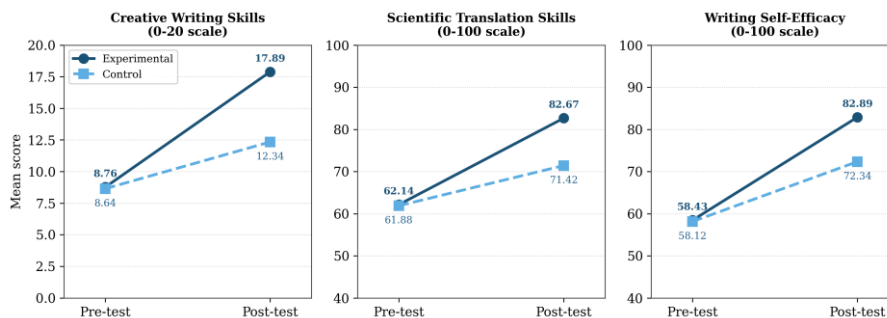


Figure 3. Within-group trajectories from pre-test to post-test for the experimental (solid lines) and control (dashed lines) groups across the three outcome measures (CWS, 0–20; STS and WSE, 0–100). The steeper experimental-group slopes illustrate the larger gains associated with automated feedback. Values correspond to Table 3.

4. Discussion

This investigation provides empirical evidence supporting the effectiveness of automated feedback systems for enhancing multiple dimensions of undergraduate writing competency. The magnitude of treatment effects—with Cohen's d values exceeding 1.80 across dependent variables—substantially surpasses typical effect sizes observed in writing intervention research.^{20,21} These findings extend previous literature examining CALL systems and automated feedback by demonstrating efficacy for sophisticated writing tasks beyond mechanical error correction.^{8,22}

4.1. Automated feedback and creative writing development

The substantial improvement in creative writing skills among the experimental group ($d = 1.84$) aligns with theoretical perspectives emphasizing the role of immediate feedback in skill acquisition and metacognitive development.^{6,23} Automated systems provided instant corrective information, enabling rapid error recognition and hypothesis testing regarding stylistic alternatives—mechanisms theoretically supported by information processing models.^{24,25}

The creative writing rubric emphasized both technical accuracy and higher-order concerns (originality, character development, narrative coherence). The substantial improvement across these dimensions suggests that automated feedback, when designed to address multiple linguistic and discourse-level features, can scaffold development of nuanced writing capabilities. The system's capacity to provide consistent feedback aligned with explicit rubric criteria likely facilitated learner understanding of quality standards and performance expectations.²⁶

Importantly, the experimental group's creative writing improvements exceeded those of the control group by a factor of approximately 3.1 (within-group d values: experimental $d_{\text{within}} = 1.56$ vs. control $d_{\text{within}} = 0.50$). This disparity suggests that traditional instructor feedback, despite its

interpretative richness, may not provide the frequency or consistency of input necessary for substantial skill development in complex writing domains.²⁷

4.2. Automated feedback and scientific translation competency

The most pronounced treatment effect emerged for scientific translation skills ($d = 2.08$), suggesting particular utility of automated systems for this specialized writing domain. Scientific translation requires simultaneous attention to terminology accuracy, semantic preservation, and stylistic conventions—cognitive demands that may particularly benefit from systematic, multi-dimensional feedback.¹²

The automated system incorporated domain-specific terminology databases and translation evaluation algorithms, potentially explaining superior outcomes compared to general writing feedback approaches. Specialized feedback addressing terminology accuracy and disciplinary conventions may provide particularly valuable input for translation learners, as these dimensions are often difficult for instructors to evaluate consistently across large student cohorts.¹²

The larger effect size for scientific translation relative to creative writing (2.08 vs. 1.84) suggests that automated systems may be particularly well-suited for structured, rule-governed tasks requiring technical precision. Conversely, creative writing—which emphasizes originality and subjective aesthetic judgment—may present greater challenges for algorithmic evaluation, potentially explaining the slightly smaller (though still substantial) effect. This interpretation is tentative and warrants direct investigation in future comparative studies.

4.3. Automated feedback and writing self-efficacy

The significant enhancement of writing self-efficacy ($d = 1.94$) provides a meaningful theoretical contribution, as self-efficacy serves as an important proximal outcome influencing persistence, engagement, and future writing endeavors.¹⁶

Theories of motivation and self-regulation emphasize the role of competence feedback and success experiences in developing confidence.^{17,28,29}

Immediate automated feedback may have facilitated frequent success experiences: students received rapid corrective input, implemented revisions, resubmitted work, and observed improvement—a cycle supporting the development of mastery experiences theorized as the most influential source of self-efficacy development.²⁸ The unlimited revision opportunities in the experimental condition enabled students to experience progressive competence development, potentially amplifying efficacy growth relative to the control condition's single-revision structure.

Notably, the writing self-efficacy improvements in the control group, while statistically significant, were substantially smaller ($d_{\text{within}} = 0.59$) than in the experimental group ($d_{\text{within}} = 1.64$). This pattern suggests that even traditional instructor feedback provides some efficacy-building benefits, but the intensity and consistency of automated feedback generate substantially more robust confidence development.

4.4. Implications for language education practice

These findings suggest that institutions considering technology integration should prioritize automated feedback system implementation, particularly for writing-intensive courses. The substantial effect sizes observed provide compelling evidence that such investments would likely yield meaningful improvements in student writing outcomes and learner confidence.

However, implementation must attend to system design features ensuring that automated feedback addresses multiple linguistic and discourse-level dimensions. Systems providing feedback only for mechanical errors or generic writing concerns may produce substantially smaller benefits than those incorporating discipline-specific and higher-order writing feedback.⁸ Professional development for faculty appears warranted, as

instructors must understand system capabilities and limitations to integrate automated feedback effectively with instructor guidance and peer collaboration.

The study also suggests potential value in combining automated and instructor feedback. Future investigations might examine hybrid approaches wherein automated systems provide rapid, frequent input while instructors focus on higher-order commentary and personalized guidance—a division of labor potentially optimizing efficiency and effectiveness.

4.5. Limitations and boundary conditions

Several limitations warrant acknowledgment. First, the study was conducted at a single institution with a relatively homogeneous sample regarding English proficiency and academic background; generalization to populations with lower proficiency or more heterogeneous experiences should proceed cautiously. Second, the automated feedback system was specifically designed for English language learners, and findings may not generalize to off-the-shelf systems lacking such specialized design. Third, although the 10-week intervention provided adequate time for skill development, longer-term follow-up would strengthen confidence in the sustainability of observed improvements. Fourth, while baseline equivalence was established statistically, unmeasured selection factors or differential motivation might have influenced outcomes. Although quasi-experimental designs cannot eliminate such threats, the effect sizes substantially exceeded those typical of selection artifacts.²⁹ Fifth, the study did not collect qualitative data to explore the mechanisms through which automated feedback influenced outcomes; future research incorporating think-aloud protocols or stimulated-recall interviews would provide mechanistic insights complementing the quantitative results.

4.6. Directions for future research

Future investigations should examine whether automated feedback efficacy persists across diverse

populations, languages, and educational contexts. Longitudinal designs tracking writing development across multiple academic years would assess the sustainability of treatment effects, and comparative studies of different system designs would clarify which features most meaningfully influence outcomes. Research examining interactions between automated feedback and individual differences (e.g., cognitive ability, motivation, writing anxiety) might identify subgroups for whom such systems are particularly or minimally beneficial.

5. Conclusion

This quasi-experimental study provides substantial empirical evidence that automated feedback systems significantly enhance undergraduate student performance across creative writing, scientific translation, and writing self-efficacy domains. The large effect sizes observed (d values: 1.84–2.08) substantially exceed typical outcomes in writing instruction research, suggesting that automated feedback represents a promising technological innovation for language education. Results support system implementation in undergraduate writing curricula, particularly when systems are designed to address multiple linguistic and discourse-level dimensions and when combined with instructor guidance and peer collaboration. As higher education increasingly emphasizes writing competency and translation skills, judicious technology integration can substantially advance student learning outcomes while providing instructors with tools for more efficient feedback provision.

6. References

1. Warschauer M. Technology and social inclusion: rethinking the digital divide. Cambridge (MA): MIT Press; 2003.
2. Chapelle CA. English language learning and technology: lectures on applied linguistics in the age of information and communication technology. Amsterdam: John Benjamins; 2003.
3. Grami GMA. The effects of integrating peer feedback into university-level ESL writing curriculum: a comparative study in a Saudi context [dissertation]. Newcastle upon Tyne: Newcastle University; 2010.
4. Kellogg RT. Training writing skills: a cognitive developmental perspective. *J Writ Res.* 2008;1(1):1-26.
5. Ellis R. Corrective feedback and teacher development. *L2 J.* 2009;1(1):3-18.
6. Nicol D. From monologue to dialogue: improving written feedback processes in mass higher education. *Assess Eval High Educ.* 2010;35(5):501-17.
7. Shermis MD, Burstein JC, editors. Automated essay scoring: a cross-disciplinary perspective. Mahwah (NJ): Lawrence Erlbaum Associates; 2003.
8. Stevenson M, Phakiti A. The effects of computer-generated feedback on the quality of writing. *Assess Writ.* 2014;19:51-65.
9. Heift T, Schulze M. Errors and intelligence in computer-assisted language learning: parsers and pedagogues. New York: Routledge; 2007.
10. Nagata N. Robo-Sensei's NLP-based error detection and feedback generation. *CALICO J.* 2009;26(3):562-79.
11. Chapelle CA, Cotos E, Lee J. Validity arguments for diagnostic assessment using automated writing evaluation. *Lang Test.* 2015;32(3):385-405.
12. Bowker L, Corpas Pastor G. Translation technology. In: Mitkov R, editor. *The Oxford handbook of computational linguistics.* 2nd ed. Oxford: Oxford University Press; 2015.
13. Morley D. *The Cambridge introduction to creative writing.* Cambridge: Cambridge University Press; 2007.
14. Flower L, Hayes JR. A cognitive process theory of writing. *Coll Compos Commun.* 1981;32(4):365-87.

15. Bereiter C, Scardamalia M. The psychology of written composition. Hillsdale (NJ): Lawrence Erlbaum Associates; 1987.
16. Pajares F. Self-efficacy beliefs, motivation, and achievement in writing: a review of the literature. *Read Writ Q.* 2003;19(2):139-58.
17. Zimmerman BJ. Becoming a self-regulated learner: an overview. *Theory Pract.* 2002;41(2):64-70.
18. Sari E, Han T. The impact of automated writing evaluation on English as a foreign language learners' writing self-efficacy, self-regulation, anxiety, and performance. *J Comput Assist Learn.* 2024;40(5):2065-80.
19. Zimmerman BJ, Bandura A. Impact of self-regulatory influences on writing course attainment. *Am Educ Res J.* 1994;31(4):845-62.
20. Lipsey MW, Wilson DB. Practical meta-analysis. Thousand Oaks (CA): SAGE Publications; 2001.
21. Graham S, McKeown D, Kiuahara S, et al. A meta-analysis of writing instruction for students in the elementary grades. *J Educ Psychol.* 2012;104(4):879-96.
22. Shermis MD. State-of-the-art automated essay scoring: competition, results, and future directions from a United States demonstration. *Assess Writ.* 2014;20:53-76.
23. Schunk DH. Self-efficacy for reading and writing: influence of modeling, goal setting, and self-evaluation. *Read Writ Q.* 2003;19(2):159-72.
24. Atkinson RC, Shiffrin RM. Human memory: a proposed system and its control processes. *Psychol Learn Motiv.* 1968;2:89-195.
25. Baddeley A. Working memory: looking back and looking forward. *Nat Rev Neurosci.* 2003;4(10):829-39.
26. Panadero E, Jonsson A. The use of scoring rubrics for formative assessment purposes revisited: a review. *Educ Res Rev.* 2013;9:129-44.
27. Ferris DR. Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In: Hyland K, Hyland F, editors. *Feedback in second language writing: contexts and issues.* Cambridge: Cambridge University Press; 2006. p. 81-104.
28. Bandura A. *Self-efficacy: the exercise of control.* New York: WH Freeman; 1997.
29. Shadish WR, Cook TD, Campbell DT. *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin; 2002.