

AI-generated visual feedback as a writing aid: Improving visual data description skills in EFL learners

Erick Purwanto^a, Xinyi Xiao^b, Junhong Liao^c, Tianchen Niu^d and Ediyanto Liu^e

^aDepartment of Computing, Xi'an Jiaotong-Liverpool University, , Erick.Purwanto@xjtlu.edu.cn; ^bXi'an-Jiaotong Liverpool University, , Xinyi.Xiao23@student.xjtlu.edu.cn; ^cXi'an-Jiaotong Liverpool University, , Junhong.Liao23@student.xjtlu.edu.cn; ^dXi'an-Jiaotong Liverpool University, , Tianchen.Niu23@student.xjtlu.edu.cn and ^eEnglish Language Centre, Xi'an Jiaotong-Liverpool University, , Ediyanto.Liu@xjtlu.edu.cn

How to cite: Purwanto, E.; Xiao, X.; Liao, J.; Niu, T.; Liu, E. (2025). AI-Generated visual feedback as a writing aid: Improving visual data description skills in EFL learners. In Y. Choubsaz, P. Díez-Arcón, A. Gimeno-Sanz, V. Morgana, A. C. Murphy & F. L. Seracini (Eds.), *Advancing CALL: New research agendas - EUROCALL 2025 Short Papers*. <https://doi.org/10.4995/EuroCALL2025.2025.21168>

Abstract

Visual feedback is essential in second language acquisition, especially for tasks that require learners to produce written descriptions of images—such as IELTS Writing Task 1—or interpret visual data in analytical writing. While recent CALL systems have begun incorporating multimodal techniques, many still rely on static images and text-based corrections, failing to address individual learners' specific errors. To address this gap, we developed VividWrite, an intelligent writing assistant that leverages an image synthesis model to produce personalized, real-time graphical representations that highlight descriptive inaccuracies while preserving the original images' visual style. Using a within-subjects design, 50 university-level EFL learners first completed a pre-intervention writing task, participated in guided sessions using VividWrite, and then completed a post-intervention assessing improvements in task accuracy, coherence, and lexical precision. Results showed significant improvements in accurately describing visual data, with notable gains in vocabulary precision. Moreover, we investigated students' perceptions of usability, writing motivation, and perceived need for teacher support. While most participants found the app beneficial, some highlighted the need for additional instructional scaffolding. These findings suggest that personalized, AI-driven visual feedback can serve as an effective pedagogical tool in CALL environments, with promising implications for broader applications across diverse linguistic tasks.

Keywords: *visual feedback; multimodal GenAI; analytical writing; intelligent writing assistant; Second Language Acquisition.*

1. Introduction

Visual corrective feedback integrates visual elements—such as color-coded highlights, graphs, or diagrams—alongside corrective textual commentary to inform learners about inaccuracies in their language use. This approach has been shown to reduce learners' cognitive load, anxiety, and boredom, while also increasing their willingness to write (Zou et al., 2025), and enhancing accuracy and motivation (Mahmood, 2024). According to Dual Coding Theory (Paivio, 1991) and Multimedia Learning Theory (Mayer & Johanson, 2008), the integration of multimodal stimuli facilitates deeper cognitive processing. Lee and Huang (2008) demonstrated that visual

aids—such as boldfacing, underlining, and color coding of target grammatical features—increase learners’ awareness of grammatical forms.

IELTS Writing Task 1 requires test-takers to produce a written description of visual data such as charts, graphs, tables, or diagrams. A common challenge is accurately identifying key trends and making appropriate comparisons. Many learners struggle with inaccurate or irrelevant information, which undermines the clarity and precision needed for high task achievement (Ahmadi et al., 2019). Effective revision involves recognizing and correcting discrepancies between writing goals and the actual text, as well as identifying suitable alternatives for unsatisfactory text (Barkaoui, 2007). In this context, describing visual data presents additional challenges, as it is unclear how to detect mismatches or evaluate alternative text effectively.

Automated Writing Evaluation (AWE) tools—such as *Grammarly*—have been shown to enhance students’ writing performance by delivering immediate, individualized feedback on grammar, punctuation, style, and mechanics (Parra & Calero, 2019). Students also reported increased motivation and engagement in the writing process. According to Link et al. (2022), AWE tools support writing development by promoting frequent and effective revisions, improving responsiveness to feedback, and enhancing grammatical accuracy with lasting retention. They also help learners internalize language rules and foster autonomy through repeated, self-paced practice. However, most AWE systems provide primarily text-based feedback and often lack support for multimodal inputs, which limit their effectiveness in tasks that require visual data interpretation.

An intelligent and interactive writing assistant is a system that autonomously generates or enhances text while actively engaging with the user’s input throughout the writing process—offering not only corrections but also supporting self-reflection that goes beyond traditional AWE tools (Lee et al., 2024). Recent tools leverage large language models (LLMs) and often output visual feedback. *Graphologue* dynamically transforms LLM responses into interactive node-link diagrams, enabling users to explore entities and relationships graphically (Jiang et al., 2023). *VISAR* supports the planning and revision of argumentative writing through an interactive hierarchical diagram that synchronizes visual and textual outlines (Zhang et al., 2023). Building on this line of research and bridging the research gaps identified above, we developed *VividWrite*, an application that produces visual feedback to highlight descriptive inaccuracies to support learners in the visual data description tasks.

2. VividWrite

We introduce *VividWrite*, an intelligent writing assistant designed to enhance second language learners’ ability to describe visual data through visual corrective feedback. VividWrite integrates an image-to-text transformer model, DePlot (Liu et al., 2023), with GPT-4o, a large language model from OpenAI, to dynamically produce graphical representations that highlight individual learners’ descriptive inaccuracies.

The process begins when the learner submits a descriptive text (see Figure 1). GPT-4o extracts linguistic features such as numbers, spatial relations, and comparisons. Simultaneously, the target image undergoes feature extraction: GPT-4o analyses color schemes and layout, while DePlot captures sequential quantitative data. These outputs are combined in the Data-driven Visualization Module. The module produces a JSON file used by the application to generate visual feedback with potential error markers, while preserving the original image’s style. In parallel, the textual data are evaluated to generate grammar, vocabulary, and coherence scores, accompanied by targeted suggestions for improvement.

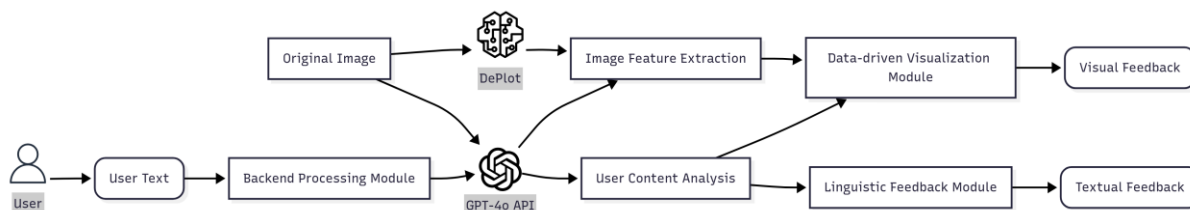


Figure 1. System architecture of VividWrite, from learner input to multimodal feedback generation.

VividWrite's interface is designed to guide learners through the writing and feedback process in a clear and intuitive manner (see Figure 2). Users begin by entering their username, which allows their progress to be tracked over time. The main writing area is positioned alongside the original target image, enabling learners to easily refer to visual data while composing their responses. The text area supports flexible editing for both brief responses and more detailed, multi-paragraph work.

VividWrite

Requirement

The bar chart below shows the total number of minutes (in billions) of telephone calls in Australia, divided into three categories, from 2001-2008. Summarise the information by selecting and reporting the main features and make comparisons where relevant. Write at least 150 words.

Original Image

Australia telephone calls by category from 2001-2008

Year	Local fixed line calls (billions)	National and international fixed line calls (billions)	Mobile calls (billions)
2001	72	38	2
2002	78	40	5
2003	82	42	10
2004	85	45	15
2005	88	48	25
2006	85	50	35
2007	80	52	40
2008	75	55	46

Student Answer

The given chart depicts the time Australian residents spent on varying types of telephone calls between 2001 and 2008.

Local fixed line calls were the highest throughout this period, upsurging from 72 billion minutes to under 90 billion in 2003. Following year, this figure peaked at 90 billion. Post this, by 2008, it had a downtrend and fell back to the figure of 2001. Both national and international fixed line calls grew gradually from 38 billion to 61 billion toward the end of the period in question. However, the progress decelerated over the last two years.

Also, dramatic growth can be seen in mobile calls from 2 billion to 46 billion minutes. This increase was specifically noticed between 2005 and 2008. During this time, the mobile phone's use got tripled. In 2008, although local fixed line calls were still popular, the gap between these three categories narrowed significantly over the second half of this period.

Student Graph

The given chart depicts the time Australian residents spent on varying types of telephone calls between 2001 and 2008.

Year	Local Fixed line (billions)	National and international fixed line (billions)	Mobile (billions)
2001	72	38	2
2002	78	40	5
2003	82	42	10
2004	85	45	15
2005	88	48	25
2006	85	50	35
2007	80	52	40
2008	75	55	46

Writing Suggestions

[Grammar Score]: 8/10
 [Grammar Issues]: The grammar is generally strong, with only minor issues. There is a slight inconsistency in verb tense usage, such as "upsurging" instead of "surged" and "got tripled" instead of "tripled." Additionally, the phrase "Post this" is informal and could be replaced with "After this."
 [Grammar Suggestions]: Ensure verb tenses are consistent and appropriate for the context. Use "surged" instead of "upsurging" and "tripled" instead of "got tripled." Replace informal phrases like "Post this" with "After this" for a more formal tone.

[Vocabulary Score]: 7/10
 [Vocabulary Issues]: The vocabulary is appropriate but somewhat limited in range. Words like "upsurging" and "got tripled" are less precise. The use of "dramatic growth" is effective, but more varied expressions could enhance the description.
 [Vocabulary Suggestions]: Expand vocabulary by using more precise terms. For example, use "increased significantly" instead of "upsurging" and "tripled" instead of "got tripled." Consider using synonyms for "growth" to add variety, such as "expansion" or "escalation."

[Coherence Score]: 9/10
 [Coherence Issues]: The composition is well-organized, with a clear progression of ideas. The flow of information is logical, and the connections between sentences and paragraphs are effective. However, the transition between discussing local fixed line calls and mobile calls could be smoother.
 [Coherence Suggestions]: To improve coherence, use transitional phrases to better connect sections. For example, "In contrast to the decline in local fixed line calls, mobile calls experienced..." This will help maintain a seamless flow between different parts of the analysis.

Figure 2. Main interface of VividWrite. (A) Username entry; (B) Task response input; (C) Button to generate visual representation of the target image description; (D) Student Graph as the visual feedback; (E) Writing suggestions with grammar, vocabulary, and coherence scores plus improvement suggestions.

Below the writing area, the "Generate Graph" button triggers the creation of a Student Graph—a visual representation of the learner's written description. By comparing this generated graph with the original image, learners can identify potential misinterpretations or omissions. This feature also helps learners compare their interpretation with the actual data trends. Alongside the graph, the system also provides scores and improvement tips for grammar, vocabulary, and coherence. Altogether, they promote self-assessment and reflection.

3. Method

3.1. Participants

The participants were 50 Chinese undergraduate university students, all of whom had prior experience with standardized English proficiency tests such as IELTS or TOEFL. Most had IELTS scores between 5.5 and 7.0 (88%), while 4% scored 5.0 or below, and 8% scored 7.5 or above. Approximately 85% were enrolled in science-related faculties, including computer science, engineering, and mathematics, while the remaining participants came from language and business programs.

3.2. Procedure

The study utilized a within-subject pre–post design. In the pre-intervention phase, participants completed an IELTS Task 1 visual data description (~ 150 words; 15–20 minutes) without feedback. During the intervention, they attended introductory sessions with multiple examples, and used VividWrite to receive visual feedback on discrepancies between their text and the target graph. Participants then revised their essays using the application's suggestions within 10–15 minutes and could regenerate visual feedback as needed. After completing these revisions, they filled out a follow-up questionnaire assessing the tool's perceived usability.

3.3. Data collection

Participants were assigned unique usernames to log into VividWrite, which stored their initial writing responses, subsequent revision attempts, and the corresponding visual feedback. At the end of the intervention, they completed an online questionnaire consisting of eight sections: demographics, writing motivation, system usability, perceived usefulness of visual feedback for revision, style similarity between the original data and generated graphs, perceived accuracy of visual feedback, and perceived need for teacher support, measured on a five-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). Two open-ended questions were also included to capture participants' subjective impressions and suggestions for improvement.

4. Results

All 50 participants completed both the pre- and post-VividWrite intervention tasks, as well as the questionnaire. Table 1 presents the results for all the questionnaire items. All six questionnaire dimensions demonstrated satisfactory internal consistency, with Cronbach's alpha values exceeding the commonly accepted threshold of 0.70, indicating reliable measurement of each construct.

Table 1. Survey statements and results.

Dimension	Statement	(n = 50)		
		M	SD	Cronbach α
Writing Motivation	I enjoyed using VividWrite to correct my essay.	3.9	0.909	0.867
	The visual feedback given by VividWrite is attractive.	4.04	1.049	
	The writing suggestions given by VividWrite are attractive.	3.84	1.017	
	I look forward to using VividWrite in the future.	4.1	1.035	
	Learning IELTS with VividWrite is interesting.	3.98	0.937	
	I actively think about what I have learned by using VividWrite.	4.32	0.741	

System Usability	VividWrite is easy to use.	4.32	0.819	0.882
	VividWrite's system performance (both visual feedback and writing suggestions) was consistent.	3.86	1.195	
	VividWrite's response time was acceptable to me.	4.08	0.944	
	The writing suggestions section was clearly displayed.	4	0.948	
	The visual feedback section in VividWrite was clearly displayed.	4.36	0.693	
	The visual feedback in VividWrite displayed the content of essay clearly	4.22	0.84	
	The visual feedback in VividWrite was easy to understand.	4.36	0.749	
Helpfulness of Visual Feedback	The visual feedback in VividWrite is useful for improving my essay. and quantitative (i.e. numeric data) descriptions.	4.18	0.873	0.861
	The visual feedback in VividWrite accurately shows the mistakes in my essay	3.92	1.066	
	The visual feedback in VividWrite helped me improve my understanding of spatial (i.e. comparisons between patterns/trends)	4.06	0.913	
	The visual feedback in VividWrite helped me organize my understanding of spatial and quantitative descriptions.	4.04	0.989	
	The visual feedback in VividWrite helped me identify where I may lose marks related to spatial and quantitative descriptions.	4.14	0.857	
	The tool supported me in organizing my essay.	3.7	1.055	
Style Similarity	The structure and color of visual feedback is like the original graph.	4.36	0.663	0.816
	The visual feedback had the same colors as the original graph.	3.84	1.131	
	The visual feedback had the same structure as the original graph.	4.12	0.961	
	If the information was correctly described in the essay, the visual feedback would be the same as the original graph.	4.16	0.738	
	The color and structure of visual feedback is well-aligned with the original graph while content of visual feedback is well-aligned with my descriptions in the essay.	3.96	0.903	
	No matter how I change my descriptions in the essay, the color and structure of visual feedback is consistent with the original graph.	3.44	1.198	
Perceived Accuracy of Visual Feedback	The visual feedback closely reflects the data or trends I discussed in my article.	4.06	0.843	0.883
	I believe the spatial layout (e.g., axes, positioning) of the visual feedback matches the descriptions in my essay.	4.22	0.648	
	The quantitative details (e.g., values, percentages) in the visual feedback accurately correspond to those in my essay.	4.36	0.749	
	VividWrite can effectively convert the content of my essay into visual feedback.	4.36	0.631	
	When describing a trend in my essays, the visual feedback represents that trend clearly.	4.2	0.728	
	Overall, I trust that the visual feedback produced is faithful visualizations of my essays content.	4.14	0.783	
Perceived Need for Teacher Support	To use VividWrite more effectively, I will need additional instructional support (e.g. a teacher's step-by-step guidance in using and understanding visual feedback).	3.54	1.014	0.745

A teacher's explanation on VividWrite's visual feedback can make it easier for me to understand and improve my essay.	3.84	0.934
In addition to feedback from VividWrite, I will still need separate feedback on my essay from a teacher.	4.02	0.795
Even after receiving all the feedback from VividWrite, I still find it difficult to write an accurate and coherent essay.	3.02	1.301
Receiving visual feedback and following a teacher's guidance will make my essay more organized and easier to understand.	4.42	0.673
When using VividWrite, I would like more help from the teacher during different steps of essay like planning, writing the first essay or revising.	4.24	0.847

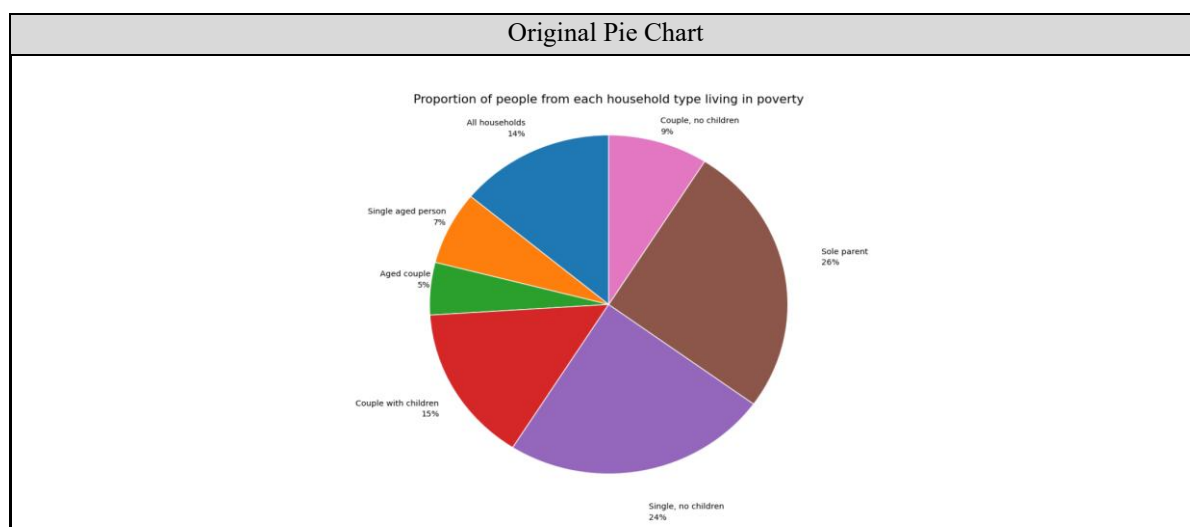
First, the Writing Motivation section—adapted from Song & Song (2023)—assessed whether VividWrite enhanced learners' interest in writing. With high internal consistency and a mean score of 4.030, results indicated that students generally agreed the tool increased their motivation. The Usefulness of Visual Feedback section (Lee et al., 2025; Ma et al., 2025; Li et al., 2025) also showed high reliability ($M = 4.007$), suggesting that targeted AI-generated visual feedback effectively supported writing improvement.

Next, the System Usability Scale (Brooke, 1995; Escalante et al., 2023) yielded a mean score of 4.171, indicating that students found VividWrite easy to use, with clear feedback and acceptable response times. Style Similarity, following the design concept of Chung et al. (2025), scored 3.980, reflecting general agreement that generated visuals matched source graphs, though perceptions of feedback stability varied. Perceived Accuracy scored highest ($M = 4.223$), with students agreeing that visual feedback accurately represented their intended content.

However, Teacher Assistance in AI-Supported Writing (Escalante et al., 2023; Awadelkarim, 2021) scored only 3.85, indicating that learners still valued an instructor's guidance. Qualitative responses further highlighted visual feedback as the tool's greatest strength, praised for making writing more engaging. Suggested improvements included clearer feedback formatting, such as using distinct background colors or highlighted modifications.

Finally, the results also present samples of learner responses showing improvements from pre- to post-intervention. These illustrate enhanced task response through more accurate visual data descriptions, greater coherence and cohesion with clearer idea progression, and richer lexical resource, including a broader vocabulary range and more precise terminology for interpreting and comparing information.

Table 2. Sample improvement in task achievement with corresponding visual feedback.



Pre-intervention	Post-intervention
<p>... Couple households showed moderate poverty levels, accounting for 15% with children and <u>14%</u> without children.</p> <p>Proportion of people from each household type living in poverty</p>	<p>... Couple households showed moderate poverty levels, accounting for 15% with children and <u>9%</u> without children. ...</p> <p>Proportion of people from each household type living in poverty</p>

In the pre-intervention task, the visual feedback indicated that 9% was missing from the student’s description. The underlined phrase highlights the error, where the student reported 14% instead of 9%. In the post-intervention task, the student corrected this mistake, and the visual feedback aligned with the original pie chart, accurately representing the key features of the task.

Table 3. Sample improvements in coherence, cohesion, and lexical resource.

Pre-intervention	Post-intervention
<p>1) ... Also, the average free hours of unemployed and retired men are about 80 hours while the women are 75 hours. Apart from that, women got approximately 40 hours free time in part-time job and 45 hours for housewives. For full-time employees, men have 8 hours more leisure time than women which reached 45 hours per week.</p> <p>2) Overall, the unemployed and retired men <u>spare</u> most leisure time, about 82 hours and 80 hours, respectively.</p> <p>3) These two types <u>consists of</u> exactly half of all poor families</p>	<p>1) ... Looking at the figures in detail, the average free hours of unemployed and retired men are about 80 hours while the women are 75 hours. <u>Notably</u>, the figures of men's leisure hours were 0 for part-time and housewives. <u>In contrast</u>, women got approximately 40 hours free time in part-time job and 45 hours for housewives. For full-time employees, men have an average of 8 hours more leisure time than women which reached 45 hours per week.</p> <p>2) Overall, the unemployed and retired men <u>allocate</u> most time <u>to</u> leisure activities, about 82 hours and 80 hours, respectively.</p> <p>3) These two types <u>comprise</u> exactly half of all poor families.</p>

In the pre-intervention task, the student primarily listed key features in a mechanical manner, relying on basic cohesive devices such as *also* and *apart from that*. In response to the writing suggestions provided by VividWrite, the student incorporated more varied connectors, such as *notably* and *in contrast*, which more clearly distinguished the data in the “part-time” and “housewives” categories, resulting in improved textual coherence. The student also enhanced lexical precision in the final draft by using the verbs *allocate* and *comprise*. However, it is worth noting that they did not adjust the tense to the past form.

5. Discussion

Findings from the Writing Motivation and System Usability dimensions indicate that VividWrite was generally perceived as engaging, easy to learn, and efficient to use. Most participants reported that the system fostered greater interest in writing. While a few items revealed wider variation in responses, suggesting that some students were less convinced of its motivational impact, the overall trend was positive. The high willingness to continue using VividWrite points to its potential for sustained engagement. Together, these results suggest that the tool supports both affective and functional aspects of writing, creating a user experience that encourages ongoing practice.

Across the Usefulness of Visual Feedback, Style Similarity, and Perceived Accuracy dimensions, participants expressed strong agreement that VividWrite's visual feedback supported their writing development. Students valued the system's ability to highlight descriptive inaccuracies, clarify quantitative and spatial relationships, and reveal potential scoring pitfalls. Perceived accuracy emerged as a notable strength, with learners trusting the visual output to reflect their intended meaning. However, the lower ratings for feedback style consistency—particularly when essay descriptions changed—point to an area for refinement. This variability suggests that while the visual feedback met most expectations, improvements to stability and graphical alignment with the original data could further enhance user trust.

Although students made measurable progress with VividWrite, results from the Teacher Assistance dimension confirm that instructional scaffolding remains important. Many participants agreed that additional guidance, such as step-by-step instructions on interpreting visual feedback, would improve their use of the system. Variation in responses suggests differing levels of self-sufficiency, likely linked to prior writing proficiency. This is consistent with the range of IELTS writing scores reported, indicating that lower-proficiency students may benefit more from supplementary teacher support. Open-ended responses reinforced this point, with several participants highlighting visual feedback as the system's greatest strength but recommending clearer formatting for writing suggestions.

Analysis of participants' pre- and post-intervention responses showed that most engaged actively with both the visual feedback and textual commentary generated by VividWrite. Several participants produced three to four revised drafts, indicating sustained interaction with the feedback process. A small number, however, did not respond to the visual feedback. Further examination revealed a pattern: some descriptions were accurate, yet VividWrite flagged data as missing. The system could not interpret expressions such as "both X and Y almost had the same ... at ... %". In these cases, participants still made use of the language-focused feedback, suggesting that even when visual feedback was less applicable, the textual component remained beneficial.

Response analysis also identified recurring challenges in IELTS Writing Task 1. The most frequent issue—missing or inaccurately described data, mirrored prior observations by Ahmadi et al. (2019)—was effectively addressed through VividWrite's visual feedback. Along with textual feedback, VividWrite supported learners struggling with lexical accuracy, such as word choice, particularly those with lower proficiency levels. Conversely, most participants demonstrated competence in identifying trends and making comparisons, likely due to prior IELTS preparation. Overall, VividWrite delivered targeted, task-specific feedback addressing diverse learner needs.

6. Conclusions

The current study was designed to assess the effectiveness of VividWrite, an AI-powered visual feedback tool, in enhancing EFL learners' ability to describe visual data accurately and coherently. The study demonstrated significant improvements in lexical precision, coherence, and task accuracy after learners engaged with personalized, multimodal feedback. Participants reported that VividWrite was engaging, easy to use, and particularly valuable for identifying and correcting descriptive inaccuracies, especially in spatial and quantitative information. Perceived accuracy of the visual feedback was a key strength of the application design.

Nevertheless, findings also revealed the continued importance of teacher support, especially for lower-proficiency learners who benefited from additional guidance in interpreting and applying feedback. Outcomes also varied by proficiency: lower-proficiency learners showed larger gains in lexical precision and error detection, whereas higher-proficiency peers leveraged the visual feedback more autonomously. While the tool's visual feedback was generally aligned with original graphs, some variability in graphical consistency indicated opportunities for refinement to enhance stability of visual outputs.

This research contributes to the growing literature on multimodal AI in CALL, demonstrating that AI-generated visual feedback can address challenges in visual data description and yield both cognitive and motivational gains. Future work should explore integrating adaptive scaffolding features, refining the alignment between feedback visuals and source data, and expanding the tool to other genres. Longitudinal studies could further examine the durability of learning gains.

Acknowledgements

The authors would like to express their sincere gratitude to all student participants for their enthusiasm, time, and valuable feedback. We also appreciate the contributions of colleagues and students who offered insightful comments during the early stages of the project, helping to refine the design of VividWrite.

References

- Ahmadi, S., Riasati, M. J., & Bavali, M. (2019). A comparison of writing performance of Iranian IELTS candidates facing chart topics vs. table topics in academic writing (Task 1). *International Journal of Instruction*, 12(4), 17–34. <https://doi.org/10.29333/iji.2019.1242a>
- Awadelkarim, A. A. (2021). An analysis and insight into the effectiveness of scaffolding: EFL instructors'/teachers' perceptions and attitudes. *Journal of Language and Linguistic Studies*, 17(2), 828–841. <https://search.informit.org/doi/10.3316/informit.216267229830245>
- Barkaoui, K. (2007). Revision in second language writing: What teachers need to know. *TESL Canada Journal*, 25(1), 81–92. <https://doi.org/10.18806/tesl.v25i1.109>
- Brooke, J. (1995). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Chung, J., Jiang, Y., Lee, S., Park, K., & Lee, B. (2025). Toyteller: AI-powered visual storytelling through toy-playing with character symbols. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM. <https://doi.org/10.1145/3706598.3713435>
- Escalante, J., Li, M., & Yang, H. H. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, Article 26. <https://doi.org/10.1186/s41239-023-00425-2>
- Jiang, P., Rayan, J., Dow, S. P., & Xia, H. (2023). Graphologue: Exploring large language model responses with interactive diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 3:1–3:20. <https://doi.org/10.1145/3586183.3606737>
- Lee, M., Gero, K. I., Chung, J. J. Y., Shum, S. B., Raheja, V., Shen, H., ... & Siangliulue, P. (2024). A design space for intelligent and interactive writing assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1-34. ACM. <https://doi.org/10.1145/3613904.3642697>

- Lee, M., Liang, P., & Yang, Q. (2022). CoAuthor: Designing a human-AI collaborative writing dataset for exploring language model capabilities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '22)*, 1–19. ACM. <https://doi.org/10.48550/arXiv.2201.06796>
- Lee, S.-K., & Huang, H.-T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, 30(3), 307–331. <https://doi.org/10.1017/S0272263108080479>
- Li, M., Wang, D., Purwanto, E., Selig, T., Zhang, Q., & Liang, H. N. (2025). VisualCodeMOOC: A course platform for algorithms and data structures integrating a conversational agent for enhanced learning through dynamic visualizations. *SoftwareX*, 30, Article 102072. <https://doi.org/10.1016/j.softx.2025.102072>
- Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605–634. <https://doi.org/10.1080/09588221.2020.1743323>
- Liu, F., Eisenschlos, J., Piccinno, F., Krichene, S., Pang, C., Lee, K., Joshi, M., Chen, W., Collier, N., & Altun, Y. (2023). DePlot: One-shot visual language reasoning by plot-to-table translation. *Findings of the Association for Computational Linguistics: ACL 2023*. <https://doi.org/10.18653/v1/2023.findings-acl.660>
- Ma, T., Al-Adwan, A. S., Li, N., Purwanto, E., Meng, W., & Liang, H.-N. (2025). Effective use of digital textbooks: A hybrid framework of acceptance and self-determination in higher education programming courses. *IEEE Transactions on Education*, 68(3), 312–321. <https://doi.org/10.1109/TE.2025.3569558>
- Mahmood, R. Q. (2024). The impact of visual corrective feedback on pronunciation accuracy in L2 sound production: Empirical evidence. In T. Bang, C. Nguyen, & H. Bui (Eds.), *Exploring contemporary English language education practices* (pp. 158–189). IGI Global. <https://doi.org/10.4018/979-8-3693-3294-8.ch008>
- Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, 63(8), 760–769. <https://doi.org/10.1037/0003-066X.63.8.760>
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45(3), 255–287. <https://doi.org/10.1037/h0084295>
- Parra, G. L., & Calero, S. X. (2019). Automated Writing Evaluation tools in the improvement of the writing skill. *International Journal of Instruction*, 12(2), 209–226. <https://doi.org/10.29333/iji.2019.12214a>
- 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, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Zhang, Z., Gao, J., Dhaliwal, R. S., & Li, T. J.-J. (2023). VISAR: A human-AI argumentative writing assistant with visual programming and rapid draft prototyping. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–30. <https://doi.org/10.1145/3586183.3606800>
- Zou, B., Wang, C., He, H., Li, C., Purwanto, E., & Wang, P. (2025). Enhancing EFL writing with visualised GenAI feedback: A cognitive affective theory of learning perspective on revision quality, emotional

response, and human–computer interaction. *Learning and Motivation*, 91, 102158.
<https://doi.org/10.1016/j.lmot.2025.102158>