

# Learners' Revision Awareness under Synthesized Image and Automated Text Feedback

Kalai WONG<sup>a\*</sup>, Yuko TOYOKAWA<sup>b</sup>, Brendan FLANAGAN<sup>b</sup> & Hiroaki OGATA<sup>b</sup>

<sup>a</sup>Graduate School of Informatics, Kyoto University, Japan

<sup>b</sup>Academic Center for Computing and Media Studies, Kyoto University, Japan

\*wong.lai.37t@st.kyoto-u.ac.jp

**Abstract:** Although many studies have highlighted the benefits of incorporating images into language learning, few have directly compared textual and visual feedback in writing contexts. Recent advances in multimodal generative AI enable the systematic generation of context-sensitive visual feedback, allowing for a more balanced comparison between feedback modalities. This study investigates differences in user perceptions and revision awareness under image-based and textual feedback conditions. A total of 67 Japanese undergraduate students participated in a writing and revision task, and 43 completed a post-task questionnaire. The results indicated no significant differences between the two groups in perceived ease of use, usefulness, enjoyment, attitude, or intention to use the system, suggesting comparable user acceptance across modalities. However, analysis of revision behaviors revealed that learners receiving image-based feedback were significantly more likely to add detailed content to their texts, whereas no significant differences were observed in surface-level revisions such as grammatical or spelling corrections. These preliminary findings suggest that while textual and visual feedback are similarly accepted by learners, image-based feedback encourages meaning-level elaboration.

**Keywords:** Image generation, EFL writing, automatic feedback, revision

## 1. Introduction

Images have become a growing focus in language learning research, particularly in vocabulary acquisition (Attygalle et al., 2025; Sugita et al., 2023; 2024). While their instructional benefits are well documented (Listyani, 2019), direct comparisons between textual and visual feedback in writing remain limited. For instance, Soria et al. (2020) combined both modalities via WhatsApp but image feedback was constrained by presettled image resources and textual feedback was more effective for detailed comments. Advances in multimodal generative AI now enable more diverse and context-sensitive visual feedback, allowing more balanced comparisons. Research has increasingly explored its role in language learning, particularly vocabulary acquisition. Prior studies suggest that generated images can support error awareness, introspection, and memory retention (Sugita et al., 2023; 2024; Attygalle et al., 2025). More broadly, images have been shown to enhance idea development, creativity, and motivation in writing (Listyani, 2019), with earlier work indicating that image creation can improve text organization and vividness (Andrzejczak et al., 2005). Recent studies integrating generative AI into writing tasks report positive learner attitudes and high enjoyment (Antony & Huang, 2025; Wong et al., 2024; 2025). However, direct comparisons between image-based and textual feedback remain limited. This study addresses this gap by examining their effects on technology acceptance and revision awareness. This study therefore examines image- versus text-based feedback in terms of user perception and revision awareness.

**RQ1.** How do users perceive the ease of use, usefulness, and enjoyment of each feedback type? What are their attitudes and intentions toward each feedback modality?

**RQ2.** How do learners self-report their revision behaviors under the two feedback conditions, and to what extent do they perceive their revisions as being influenced by the provided feedback?

## 2. Method

Data were collected during a university lecture in Japan with 67 undergraduates randomly assigned to image- or textual-feedback groups. Image feedback was generated using Gemini 2.5 Flash, while textual feedback was produced with GPT-4o mini using a structured prompt (grammar, spelling, style, clarity, and evaluation). Participants completed a descriptive writing task, revised their work after feedback, and a 5-minute voluntary questionnaire; 43 responses were obtained (image:  $n = 16$ ; textual:  $n = 27$ ). User perception was measured using the exTAM questionnaire (Wu & Gao, 2011), assessing PEOU, PU, PE, ATT, and intention on a 7-point Likert scale. Revision awareness was examined via two multiple-choice items, reported vs. feedback-induced revisions, (Faigley & Witte, 1981; Zhang & Hyland, 2023) and one open-ended question. Based on prior taxonomies, revision types included grammar, spelling, lexical changes, paraphrasing, and content addition. Fisher's exact tests were used due to small expected cell counts.

## 3. Result

ExTAM results indicated generally positive perceptions in both groups. The textual-feedback group showed slightly higher scores for PEOU, PU, and intention, whereas the image-feedback group reported marginally higher enjoyment and attitude. Mann–Whitney U tests, conducted due to non-normality and unequal group sizes, found no significant differences across constructs (all  $p > .05$ ), suggesting comparable acceptance of both feedback types.

Table 1. Descriptive statistics of *exTAM result*

	Image Feedback ( $n=16$ )			Textual Feedback ( $n=27$ )		
	Mean	STD	Cronbach's $\alpha$	Mean	STD	Cronbach's $\alpha$
PEOU	4.75	1.85	.97	<b>5.05</b>	1.47	.96
PU	4.98	1.49	.91	<b>5.04</b>	1.62	.96
PE	<b>5.23</b>	2.09	.97	5.16	1.92	.99
ATT	<b>4.75</b>	1.62	.75	4.69	1.33	.43*
Intention	4.15	1.94	.97	<b>4.37</b>	1.75	.97

\* ATT reliability was initially low ( $\alpha = .43$ ) due to straightlining and a misaligned reverse-coded item; excluding five cases increased reliability to  $\alpha = .77$  ( $M = 4.64$ ,  $SD = 1.47$ ;  $n = 22$ ).

Fisher's exact tests examined associations between feedback type and revision awareness. No significant relationships were found for grammatical, spelling, or lexical corrections ( $ps \geq .52$ ). However, a significant association emerged for adding content details, with learners in the image-feedback group more likely to elaborate than those in the textual group ( $p = .002$ ). A similar pattern was observed for feedback-induced revisions: no significant effects for grammar or spelling ( $ps \geq .28$ ), but a significant effect for content addition ( $p = .004$ ). Paraphrasing was not analyzed due to no occurrences. Overall, feedback modality was unrelated to surface-level revisions but was associated with greater content elaboration in the image-feedback condition.

Table 2. Frequency table of *revision awareness and feedback induced revision* ( $n=43$ )

	Revision effort			Feedback induced revision		
	Image	Textual	$p$	Image	Textual	$p$

Grammatical correction	1	3	>.99	1	5	.39
Spelling correction	0	2	.52	0	4	.28
Adding content detail	7	1	.002	9	3	.004
Lexical correction	0	1	>.99	0	0	-
Paraphrasing	0	0	-	0	0	-

#### 4. Discussion and Conclusion

The findings provide exploratory evidence that both image and textual feedback were positively received, with no significant differences in technology acceptance measures. Minor variations suggest that image feedback may enhance engagement, while textual feedback may be perceived as more useful, potentially influencing intention. Qualitative responses highlight complementary affordances: textual feedback supports explicit correction, whereas image feedback may encourage reflection and content elaboration. Although no differences were found in surface-level revisions, image feedback was associated with increased detail addition, suggesting a possible shift toward meaning-level revision. Given limitations including small sample size, self-selection bias, brief intervention, and reliance on self-reports, these results should be interpreted cautiously as preliminary trends. Further research is required to confirm and extend these findings.

#### Acknowledgements

This work was supported in part by the following grants: (B) JP23H01001, (A) JP23H00505 and New Energy and Industrial Technology Development Organization (NEDO) under Grant JPNP20006 and Grant JPNP18013, and CSTI 3rd SIP Program JPJ012347.

#### References

- Andrzejczak, N., Trainin, G., & Poldberg, M. (2005). From Image to Text: Using Images in the Writing Process. *International Journal of Education & the Arts*, 6(12), 1-17.
- Antony, V. N., & Huang, C. M. (2025). ID. 8: Co-creating visual stories with generative AI. *ACM Transactions on Interactive Intelligent Systems*, 14(3), 1-29.
- Attygalle, N. T., Kljun, M., Quigley, A., Čopič Pucihar, K., Grubert, J., Biener, V., ... & Weerasinghe, M. (2025, March). Text-to-image generation for vocabulary learning using the keyword method. In *Proceedings of the 30th International Conference on Intelligent User Interfaces* (pp. 1381-1397).
- Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition & Communication*, 32(4), 400-414.
- Listyani, L. (2019). The Use of a Visual Image to Promote Narrative Writing Ability and Creativity. *Eurasian Journal of Educational Research*, 80, 193-223.
- Soria, S., Gutiérrez-Colón, M., & Frumuselu, A. D. (2020). Feedback and Mobile Instant Messaging: Using WhatsApp as a Feedback Tool in EFL. *International Journal of Instruction*, 13(1), 797-812.
- Sugita, K., Ota, K., Gu, W., & Hasegawa, S. (2023). *L-VEIGe: Vocabulary Learning Support System Using Error Image Generation A study of criteria for image suggestibility*. JSiSE Research Report, 38 (2), 118-124.
- Sugita, K., Gu, W., Ota, K., & Hasegawa, S. (2024). L-VEIGe: Development of L2 Vocabulary Learning Support System with Suggestibility of Error by Image Generation.
- Wong K., Ocheja, P., Flanagan, B., & Ogata, H. (2024, November). AVERY: A GenAI-Based Approach to Enhancing Learner Engagement in English Writing. In *International Conference on Computers in Education*.
- Wong, K., Toyokawa, Y., Dai, Y., Flanagan, B., & Ogata, H. (2025, September). Picture-Cued Writing Using AI-Generated Images for Language Acquisition. In *International Conference on Learning Evidence and Analytics*.
- Wu, X., & Gao, Y. (2011). Applying the Extended Technology Acceptance Model to the Use of Clickers in Student Learning: Some Evidence from Macroeconomics Classes. *American Journal of Business Education*, 4(7), 43-50.
- Zhang, Z. V., & Hyland, K. (2023). Student engagement with peer feedback in L2 writing: Insights from reflective journaling and revising practices. *Assessing Writing*, 58, 100784.