

AI-Assisted Writing Feedback in EFL: Tracking Student Performance and Reflections

Junko Otoshi
Naomi Fujishima

JALTCALL 2025
July 19 2025

Outline of Today's Presentation

- Purpose of Study
- What is Write & Improve?
- Theoretical Orientation
- Action Research
- Results
- Discussion and Summary
- Implications

Purpose of Study

This study aims to explore how EFL university students develop their writing proficiency and feedback literacy through engagement with AI-assisted feedback from Write & Improve.

What is Write & Improve?

Write & Improve
with Cambridge



Feedback Screen

Write & Improve

with Cambridge

CEFR
Score



Source : Cambridge University Press & Assessment

Word Feedback

Hi Chris, how are you? I'm going to a party of my friend of school. He's name me? It begins at 7 o'clock and it will have music and we will dance. It is in his house. He lives → in ← same street as me. You can come here at 7 o'clock. Please ! to come! Mary

Missing word?

Did you forget something after this word? Do you need to add 'the' or something similar?



Getting started with Write & Improve by English with Cambridge

Theoretical Orientation

- In a sociocognitive perspective, engagement with feedback is not limited to the act of revising texts but involves cognitive, behavioral, emotional, and agentic dimensions (Mao & Lee, 2022).
- Feedback literacy is an ability to interpret information and use it to improve tasks or learning processes (Suttons, 2012).

Theoretical Orientation

- In a sociocognitive perspective, engagement with feedback is not limited to the act of revising texts but involves cognitive, behavioral, emotional, and agentic dimensions (Mao & Lee, 2022).
- Feedback literacy is an ability to interpret information and use it to improve tasks or learning processes (Suttons, 2012).

Action Research

- Track students' writing development using an AI-assisted writing practice tool
- Foster students' feedback literacy through self-reflection logs

Term 1 (7 sessions)

Term 2 (7 sessions)

A median split was applied to create balanced performance-based groups (Tabachnick & Fidell, 2019).

21 students in EFL academic writing class (9=upper level; 12=lower level)

Term 1 (7 sessions)

Term 2 (7 sessions)

21 students in EFL academic writing class (9=upper level; 12=lower level)

Independent Study

Platform: **Write & Improve (W&I)**

(**5** tasks in Term 1; **4** tasks in Term 2)

Term 1 (7 sessions)

Term 2 (7 sessions)

21 students in EFL academic writing class (9=upper level; 12=lower level)

Instructor's role:

designed task prompts and provided minimal feedback with some emotional feedback (e.g, praises) for each task.

Term 1 (7 sessions)

Term 2 (7 sessions)

21 students in EFL academic writing class (9=upper level; 12=lower level)

Student's role:

submitted writing to W&I;
reviewed CEFR scores and automated feedback;
completed **a self-reflection log** for each task

Term 1 (7 sessions)

Term 2 (7 sessions)

21 students in EFL academic writing class (9=upper level; 12=lower level)

Independent Study

Platform: **Write & Improve (W&I)**

(5 tasks in Term 1; 4 tasks in Term 2)

Instructor's role:

designed task prompts and provided minimal feedback

Student's role

submitted writing to W&I;

reviewed CEFR scores and automated feedback

completed a self-reflection log for each task

In-class Exam (280–300 words)
Narrative essay assessed **by W&I**

Take-home final essay (350 words)
Argumentative essay assessed
by **2 human teachers**

Research Questions

1. How do students' writing scores using Write & Improve change across tasks during the course? What patterns of development emerge between higher- and lower-level writing groups?

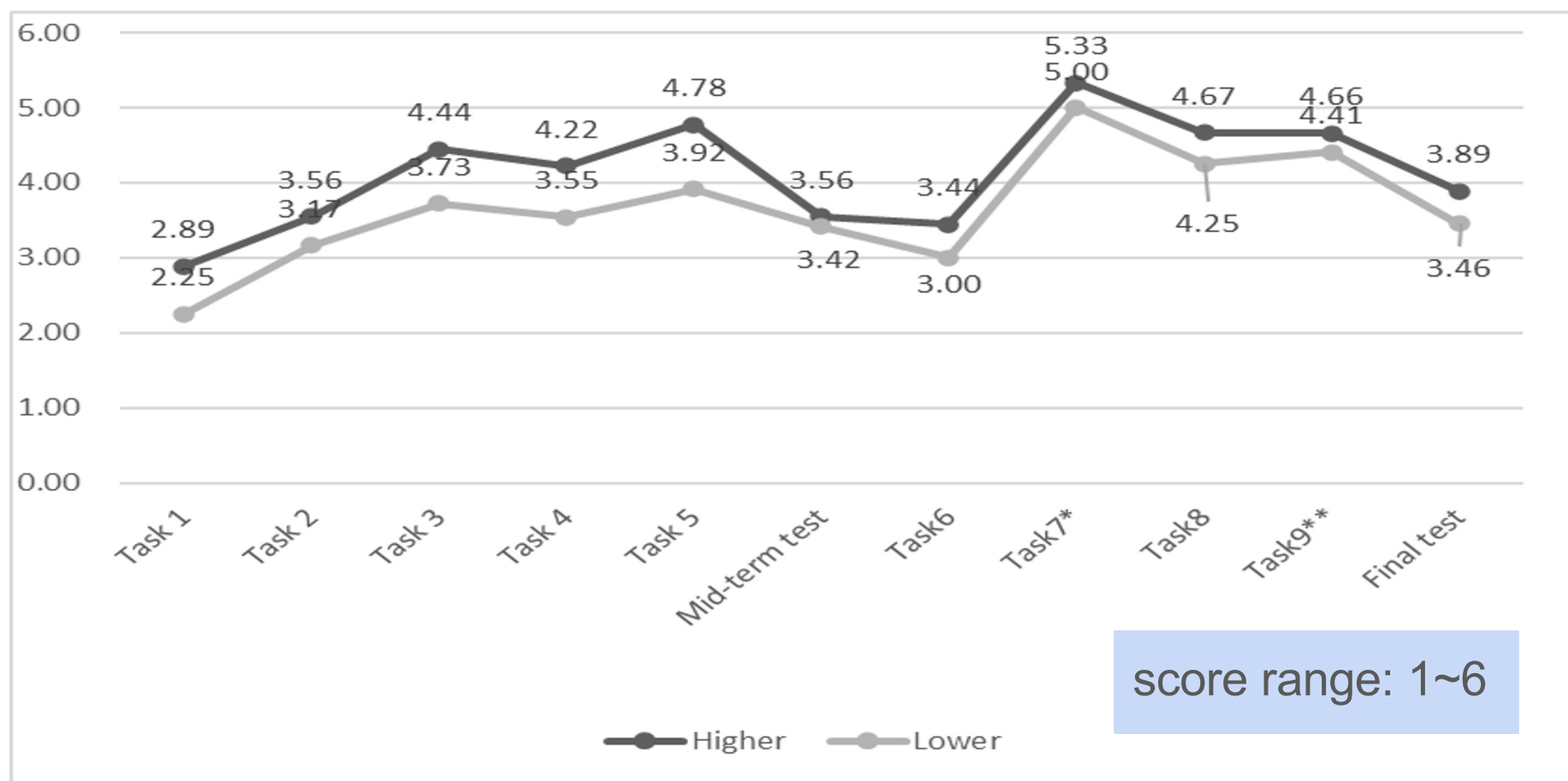
Research Questions

2. How do students' self-reflection comments reveal their engagement with AI feedback, their feedback literacy, and the challenges they experience over time?

Results

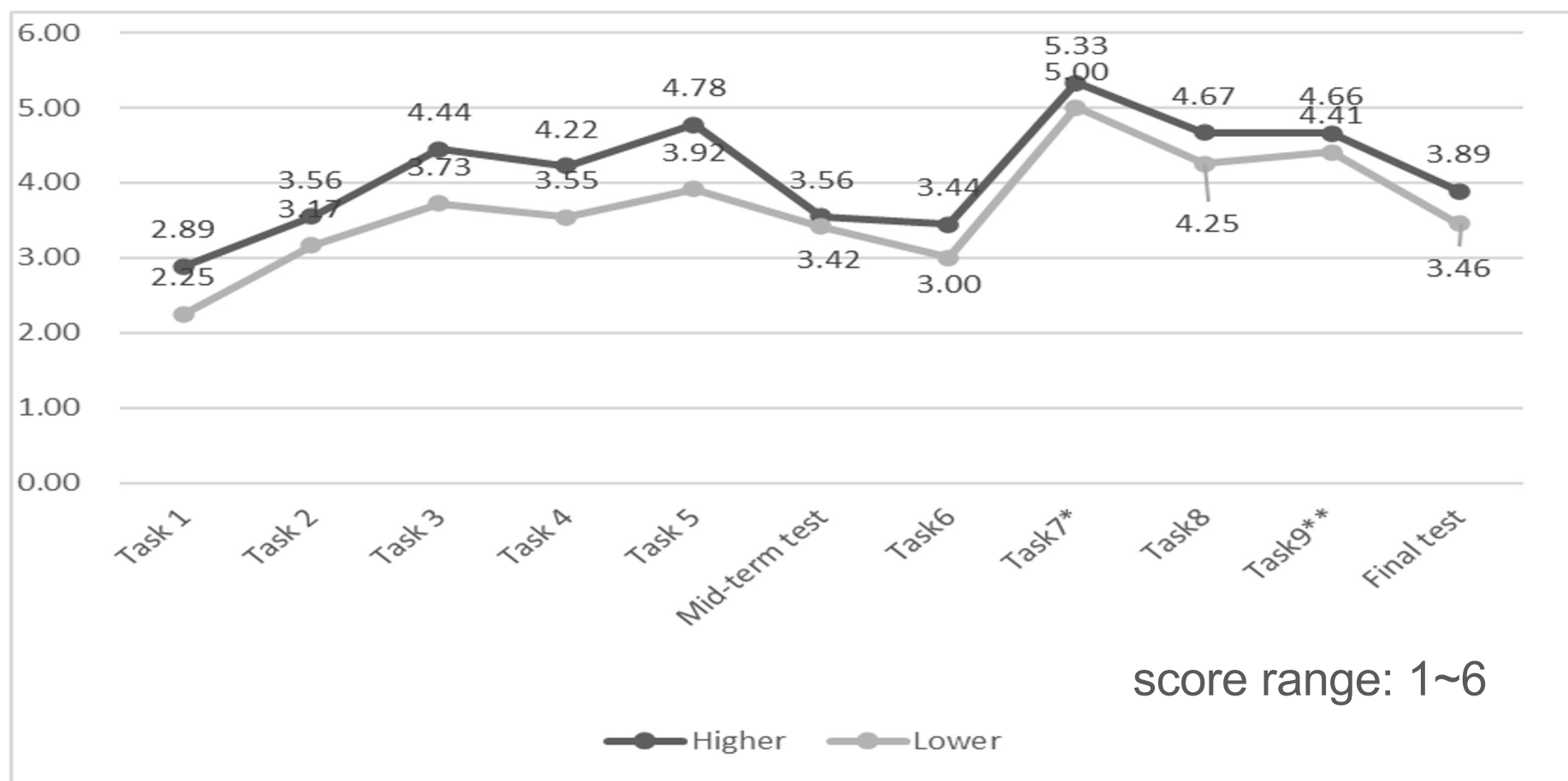
1. How do students' writing scores using Write & Improve change across tasks during the course? What patterns of development emerge between higher- and lower-level writing groups?

Transition of Average Writing Task Scores between Groups



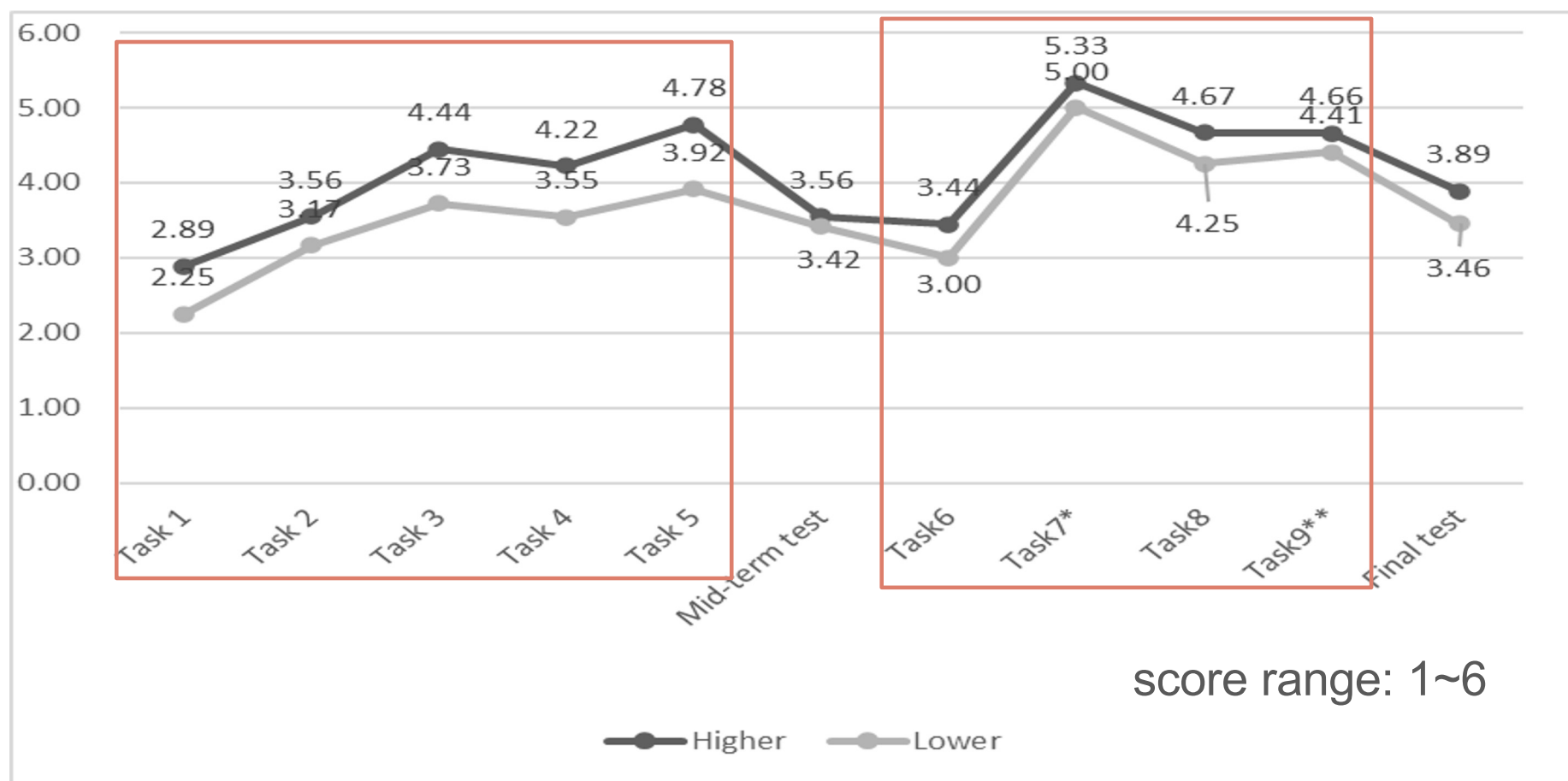
Note: CEFR levels are converted to numerical values: A1=1, A2=2, B1=3, B2=4, C1=5, C2=6.
Final test was evaluated by two human teachers.

Transition of Average Writing Task Scores between Groups



Note: CEFR levels are converted to numerical values: A1=1, A2=2, B1=3, B2=4, C1=5, **C2=6**.
Final test was evaluated by two human teachers.

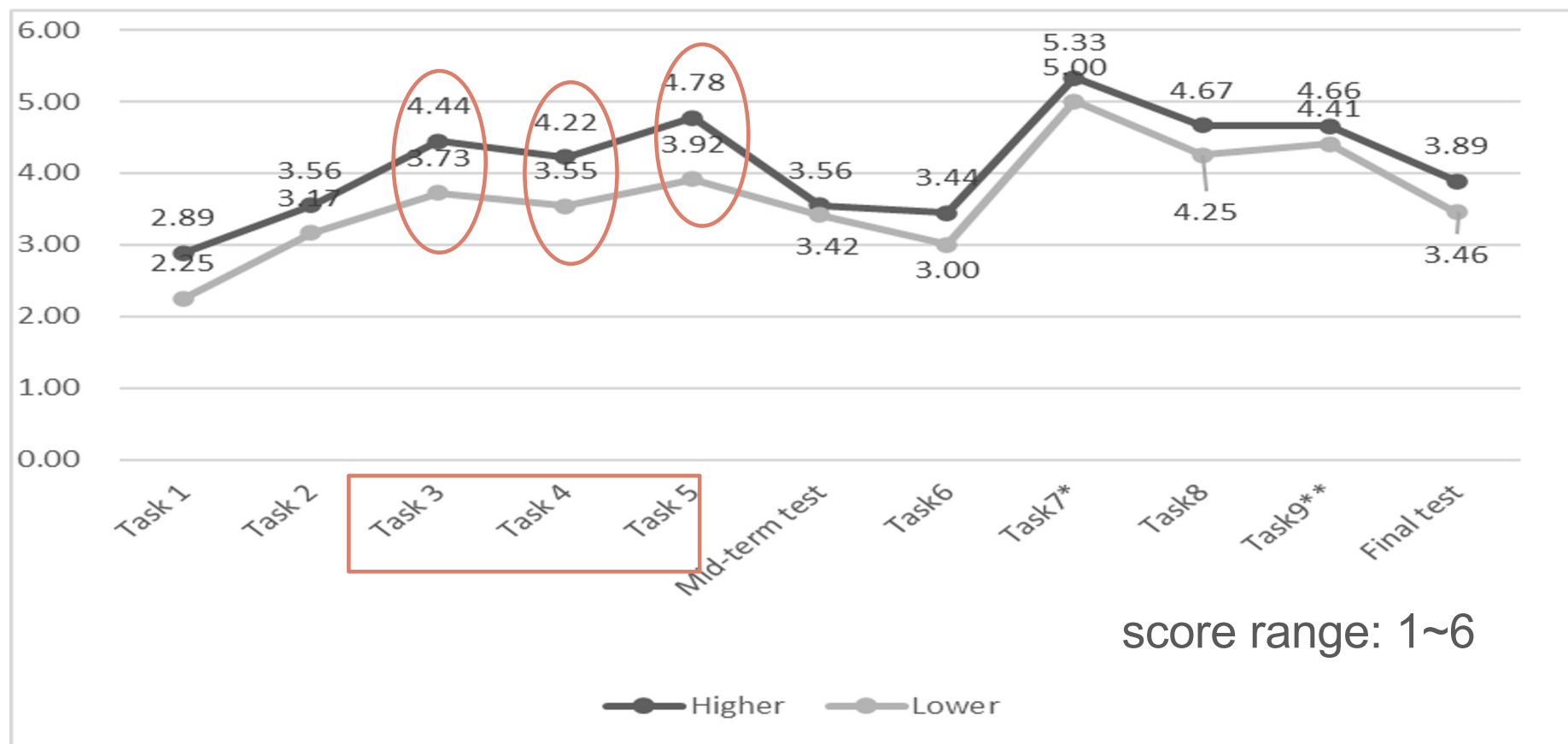
Transition of Average Writing Task Scores between Groups



Note: CEFR levels are converted to numerical values: A1=1, A2=2, B1=3, B2=4, C1=5, C2=6.

Final test was evaluated by two human teachers.

Transition of Average Writing Task Scores between Groups



score range: 1~6

Tasks 3–5 are descriptive/narrative essays (approx. 300 words).

Tasks 6–9 are argumentative essays (approx. 350 words).

Mann-Whitney U test was used to examine group differences.

1. How do students' writing scores using Write & Improve change across tasks during the course? What patterns of development emerge between higher- and lower-level writing groups?

- ❑ The result suggests that the performance gap was most pronounced in the early-to-mid stages of the course but narrowed as the course progressed.
- ❑ Both groups demonstrated developmental gains across linguistic features; however, the lower group exhibited substantial improvements, particularly in complexity, fluency, and sophistication (measured by CEFR-based Writing Level Analyzer, Uchida & Negishi, 2025).

1. How do students' writing scores using Write & Improve change across tasks during the course? What patterns of development emerge between higher- and lower-level writing groups?

- ❑ The result suggests that the performance gap was most pronounced in the early-to-mid stages of the course but narrowed as the course progressed.
- ❑ Both groups demonstrated developmental gains across linguistic features; however, the lower group exhibited substantial improvements, particularly in complexity, fluency, and sophistication (measured by CEFR-based Writing Level Analyzer, Uchida & Negishi, 2025).

2. How do students' self-reflection comments reveal their engagement with AI feedback, their feedback literacy, and the challenges they experience over time?

Guided questions to facilitate the self-reflection comments for every task:

1. Do you think the score you received accurately reflects the quality of your essay?
1. What did you learn from the W&I feedback? Was any specific feedback particularly useful, or was any part unclear? Please provide a brief comment on your experience.

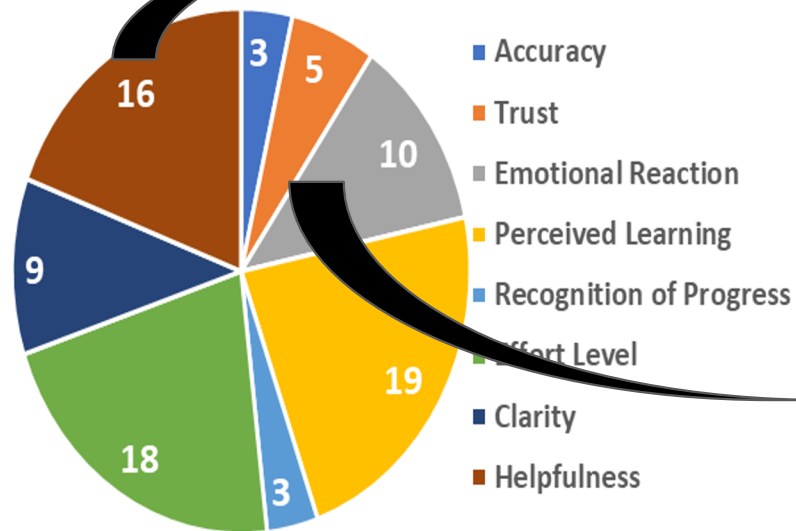
Category	Focused categories	Definition
AI Assessment	Accuracy	Comments on the perceived accuracy or fairness of AI feedback.
	Trust	Expressions of trust or skepticism towards the assessment by AI.
Personal Experience	Emotional Reaction	Comments reflecting emotions such as frustration, satisfaction, or confidence after receiving feedback.
	Perceived Learning	Statements about what students felt they learned or struggled with after completing tasks.
	Recognition of Progress	Reflections on perceived improvement in writing ability.
	Effort Level	Descriptions of the effort students intended to put into future tasks.
AI Feedback Utility	Clarity	Comments on the clarity of feedback and any confusion encountered.
	Helpfulness	Comments on useful feedback, especially on linguistic aspects like vocabulary, grammar, or spelling.

Category	Focused categories	Definition
AI Assessment	Accuracy	Comments on the perceived accuracy or fairness of AI feedback.
	Trust	Expressions of trust or skepticism towards the assessment by AI.
Personal Experience	Emotional Reaction	Comments reflecting emotions such as frustration, satisfaction, or confidence after receiving feedback.
	Perceived Learning	Statements about what students felt they learned or struggled with after completing tasks.
	Recognition of Progress	Reflections on perceived improvement in writing ability.
	Effort Level	Descriptions of the effort students intended to put into future tasks.
AI Feedback Utility	Clarity	Comments on the clarity of feedback and any confusion encountered.
	Helpfulness	Comments on useful feedback, especially on linguistic aspects like vocabulary, grammar, or spelling.

Category	Focused categories	Definition
AI Assessment	Accuracy	Comments on the perceived accuracy or fairness of AI feedback.
	Trust	Expressions of trust or skepticism towards the assessment by AI.
Personal Experience	Emotional Reaction	Comments reflecting emotions such as frustration, satisfaction, or confidence after receiving feedback.
	Perceived Learning	Statements about what students felt they learned or struggled with after completing tasks.
	Recognition of Progress	Reflections on perceived improvement in writing ability.
	Effort Level	Descriptions of the effort students intended to put into future tasks.
AI Feedback Utility	Clarity	Comments on the clarity of feedback and any confusion encountered.
	Helpfulness	Comments on useful feedback, especially on linguistic aspects like vocabulary, grammar, or spelling.

Lower Group's Response Changes Analyzed by Fisher's Exact Test

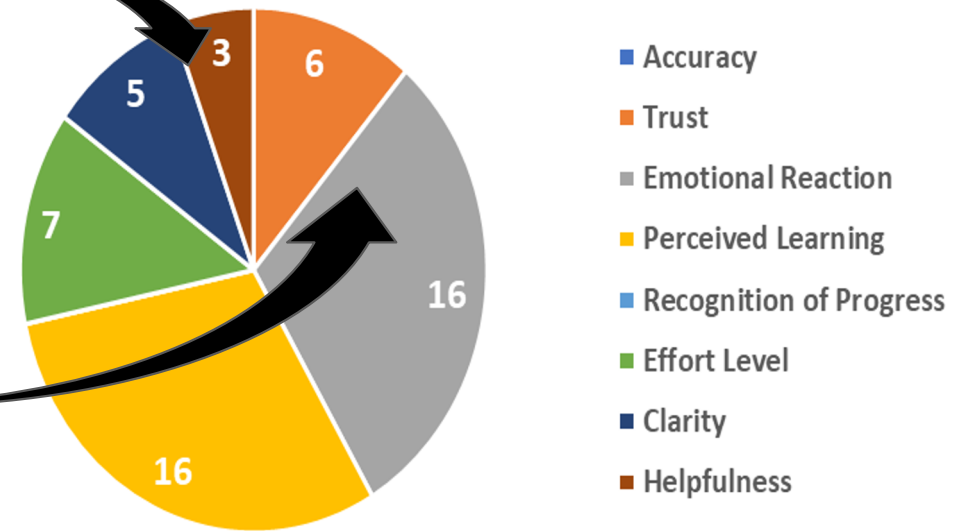
Helpfulness: 19.2% → 5.6%



Term 1

83 counts for 5 tasks

Emotional Reaction: 12% → 30%



Term 2

53 counts for 4 tasks

Self-Reflection Logs by the Lower Group Students on Emotional Reaction

Term 1	Term 2
<ul style="list-style-type: none">● I think it improved my writing skill little. It was interesting.● It was very difficult to make my own story, but I enjoyed it a little.	<ul style="list-style-type: none">● The CEFR level decreased. I was shocked about it.● I did my best, but my level did not rise.● I revised my essay several times, but my scores did not improve (I used "Chat GPT", too)

Self-Reflection Logs by the Lower Group Students on Helpfulness

Term 1	Term 2
<ul style="list-style-type: none">● And I made some mistakes of spelling.● I still made a lot of mistakes about sentence, but I found where mistake is.	<ul style="list-style-type: none">● I made a lot of mistakes of spelling.● I saw <i>Change</i> that was made by Write & Improve. I am surprised to see it because there are a lot of mistakes.

Perceived Learning:

Perceived Learning: The Largest of the Focused Categories

Operational Definition:

Statements about what students felt they learned or struggled with after completing tasks.

No statistically significant changes were observed between terms; however, the content of the comments became more metacognitive.

Perceived Learning:

Perceived Learning: The Largest of the Focused Categories

Operational Definition:

Statements about what students felt they learned or struggled with after completing tasks.

No statistically significant changes were observed between terms; however, the content of the comments became more metacognitive.

Perceived Learning:

Perceived Learning: The Largest of the Focused Categories

Defined as:

Statements about what students felt they learned or struggled with after completing tasks.

No statistically significant changes were observed between terms; however, the content of the comments became more metacognitive.

Lower Group Self-Reflection Logs on Perceived Learning

Term 1	Term 2
<ul style="list-style-type: none">● I made a lot of mistakes in spelling and grammar.● I tried to use not ordinary adjectives, but it may not fit for my essay.	<ul style="list-style-type: none">● I may have written things that don't fit the theme.● I used specific examples in this writing. I think it was very difficult to summarize the content using that specific example.● ...but it was difficult to write a correct assertion.

2. How do students' self-reflection comments reveal their engagement with AI feedback, their feedback literacy, and the challenges they experience over time?

Initial Stages:

AI was helpful for surface-level corrections, especially for lower-level students.

Changes Over Time:

- **Emotional Reaction:** Became more negative
- **Helpfulness:** Decreased

Feedback Literacy:

Students became more metacognitively aware, increasingly focusing on rhetorical and logical aspects.

2. How do students' self-reflection comments reveal their engagement with AI feedback, their feedback literacy, and the challenges they experience over time?

Initial Stages:

AI was helpful for surface-level corrections, especially for lower-level students.

Changes Over Time:

- **Emotional Reaction:** Became more negative
- **Helpfulness:** Decreased

Feedback Literacy:

Students became more metacognitively aware, increasingly focusing on rhetorical and logical aspects.

2. How do students' self-reflection comments reveal their engagement with AI feedback, their feedback literacy, and the challenges they experience over time?

Initial Stages:

AI was helpful for surface-level corrections, especially for lower-level students.

Changes Over Time:

- **Emotional Reaction:** Became more negative
- **Helpfulness:** Decreased

Feedback Literacy:

Students became more metacognitively aware, increasingly focusing on rhetorical and logical aspects.

Discussion and Summary

- **W&I and Writing Development**

Independent writing practice using *Write & Improve* (W&I) facilitated students' writing development process as task complexity increased.

- **Greater Gains in Lower-Level Group**

This trend was particularly noticeable among lower-scoring students, who showed gradual improvements in **complexity**, **fluency**, and **sophistication**.

- **Frustration Yet Growth**

Despite expressing frustration over a perceived lack of progress and limited support from W&I, these students gradually developed a **deeper metacognitive awareness** of genre-specific writing conventions and linguistic features.

Discussion and Summary

- **W&I and Writing Development**

Independent writing practice using *Write & Improve* (W&I) supported students' writing development as task complexity increased.

- **Greater Gains in Lower-Level Group**

This trend was particularly noticeable among lower-scoring students, who showed gradual improvements in complexity, fluency, and sophistication.

- **Frustration Yet Growth**

Despite expressing frustration over a perceived lack of progress and limited support from W&I, these students gradually developed a **deeper metacognitive awareness** of genre-specific writing conventions and linguistic features.

Discussion and Summary

- **W&I and Writing Development**

Independent writing practice using *Write & Improve* (W&I) supported students' writing development as task complexity increased.

- **Greater Gains in Lower-Level Group**

This trend was particularly noticeable among lower-scoring students, who showed gradual improvements in complexity, fluency, and sophistication.

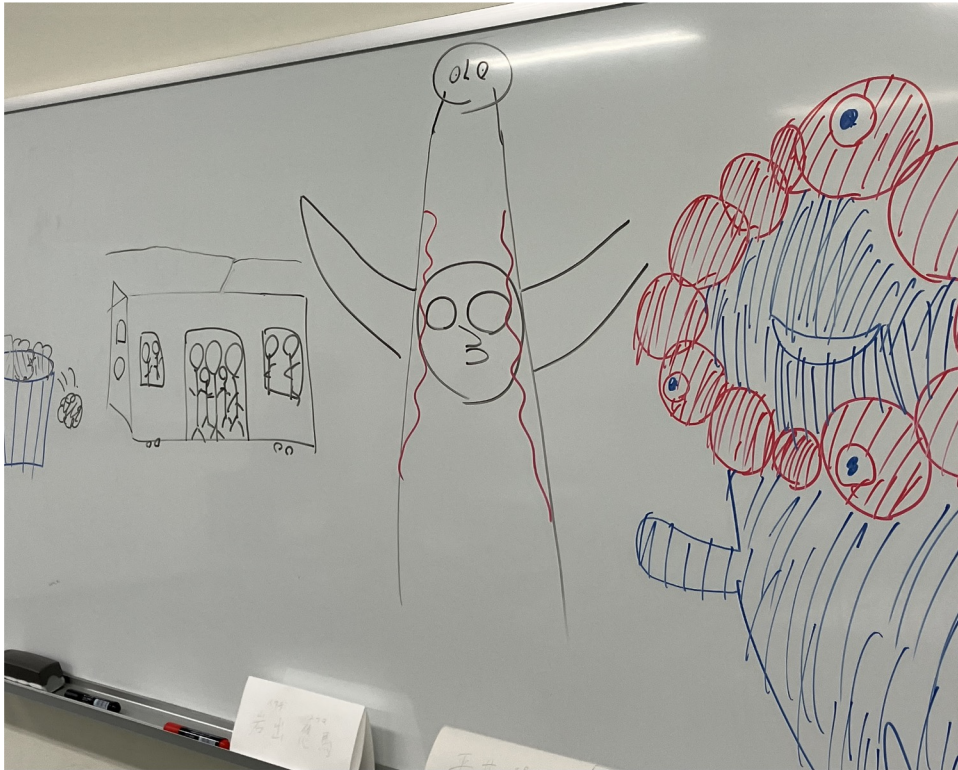
- **Frustration Yet Growth**

Despite expressing frustration over a perceived lack of progress and limited support from W&I, self-reflection comments showed that these students gradually developed a **deeper metacognitive awareness**, demonstrating interpretive skills associated with feedback literacy.

Implications from the Study: Human in the Loop Approach

- **Generative AI** that delivers guided, sentence-level revisions with explanations may help students refine their writing skills, beyond what surface-level correction tools offer.
- **Teachers** should help students develop the feedback literacy needed to avoid blind trust in AI-generated feedback and to complement it with personalized, constructive comments that support effective application to future tasks.
- **Human interaction** such as peer-review activities to better understand how their writing is received by real readers—especially to develop an audience awareness.

Do you want to hand over your humanity to AI?



Thank you!

Any Questions?

References

- Anson, C. M., Dannels, D. P., Laboy, J., & Carneiro, S. (2021). Students' perceptions of the affordances of teacher and AI feedback on writing. *Assessing Writing*, 49, 100520. <https://doi.org/10.1016/j.asw.2021.100520>
- Burnell, C., Lin, C. Y., Mehta, S., Shah, M., & Wu, Y. (2023). Reflective practices in AI-assisted learning: A formative assessment perspective. *Journal of Educational Technology & Society*, 26(1), 45–60.
- Cambridge English. (2024). *Write & Improve + Class View*. <https://writeandimprove.com/>
- Ferris, D. R., & Hedgcock, J. S. (2014). *Teaching L2 composition: Purpose, process, and practice* (3rd ed.). Routledge.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*. Longman.
- Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, 30, 31–44. <https://doi.org/10.1016/j.jslw.2015.08.002>
- Hyland, F., & Hyland, K. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101. <https://doi.org/10.1017/S0261444806003399>
- Jia, J., Xu, H., Chen, C., & Wang, L. (2023). The impact of AI-generated feedback on L2 writing: A meta-analysis. *System*, 114, 102964. <https://doi.org/10.1016/j.system.2023.102964>
- Kao, C.-W., & Reynolds, B. L. (2024). Timed second language writing performance: Effects of perceived teacher vs perceived automated feedback. *Humanities and Social Sciences Communications*, 11, Article 292. <https://doi.org/10.1057/s41599-024-03522-3>
- Kasneci, E., Sessler, K., & Seegerer, S. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Lau, J., Chan, T., & Tong, A. (2024). Learner engagement with feedback in automated writing evaluation. *Assessing Writing*, 61, 100752. <https://doi.org/10.1016/j.asw.2024.100752>

Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Springer.

Lee, I. (2019). Engaging students in learning to write: Feedback in the writing classroom. In X. Gao (Ed.), *Second Handbook of English Language Teaching* (pp. 1–17). Springer. https://doi.org/10.1007/978-3-030-02899-2_19-1

Lee, I., Mak, P., & Yuan, R. E. (2019). Assessment as learning in writing: A review. *Assessing Writing*, 39, 14–25. <https://doi.org/10.1016/j.asw.2018.12.004>

Li, Z., Link, S., & Hegelheimer, V. (2015). Rethinking teacher roles in writing instruction with technology. *Journal of Second Language Writing*, 27, 1–17. <https://doi.org/10.1016/j.jslw.2014.10.004>

Mao, Z., & Lee, I. (2022). Feedback engagement in L2 writing: A dynamic perspective. *Language Teaching Research*, 26(5), 832–857. <https://doi.org/10.1177/1362168820933192>

McDonald, J. H. (2014). *Handbook of biological statistics* (3rd ed.). Sparky House Publishing.

Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, 15(2), 118–141. <https://doi.org/10.1016/j.jslw.2006.01.003>

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Language Testing in Asia*, 13, Article 32. <https://doi.org/10.1186/s40468-023-00230-7>

Reid, J. M. (1993). *Teaching ESL writing*. Prentice Hall Regents.

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). Sage.

Shi, L., & Aryadoust, V. (2024). Student engagement with automated writing evaluation: A mixed-methods study. *Language Testing*, 41(2), 153–176.
<https://doi.org/10.1177/02655322231211233>

Shin, D., & Lee, J. H. (2024). Exploratory study on the potential of ChatGPT as a rater of second-language writing. *Education & Information Technologies*, 29, 24735–24757.
<https://doi.org/10.1007/s10639-024-13061-5>

Shvidko, E., Evans, N. W., & Donahue, C. (2020). The role of emotion in written corrective feedback: Exploring teacher and student perspectives. *TESOL Quarterly*, 54(3), 656–685. <https://doi.org/10.1002/tesq.566>

Sutton, P. (2012). Conceptualizing feedback literacy: Knowing, being, and acting. *Innovations in Education and Teaching International*, 49(1), 31–40.
<https://doi.org/10.1080/14703297.2012.647781>

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Uchida, A., & Negishi, M. (2025). CEFR-based writing level analyzer version 2.1 [Computer software]. CEFR-J Project.

van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Springer.

Wang, W. (2023). Exploring students' perceptions of automated writing feedback in EFL contexts. *Computer Assisted Language Learning*, 36(1–2), 129–152.
<https://doi.org/10.1080/09588221.2021.1938450>

Wang, Y., & Lee, I. (2021). Student engagement with feedback: Beyond the cognitive dimension. *Assessing Writing*, 47, 100511.
<https://doi.org/10.1016/j.asw.2020.100511>

Yang, Y., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, 15(3), 179–200. <https://doi.org/10.1016/j.jslw.2006.09.004>

Yang, W., Wang, X., & Liu, L. (2024). Understanding EFL learners' uptake of automated feedback: The role of reflection. *Language Learning & Technology*, 28(1), 1–22. <https://doi.org/10.1016/j.chb.2023.107720>

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, 49(4), 461–493. <https://doi.org/10.1017/S0261444816000161>

Zemach, D. E., & Rumisek, L. A. (2005). *Academic writing: From paragraph to essay*. Macmillan.

Zheng, B. (2021). Emerging technologies for second language writing instruction: The case of AI. *Computer Assisted Language Learning*, 34(3), 208–231. <https://doi.org/10.1080/09588221.2020.1729837>

Zhong, Q., & Yang, L. (2021). Assessment as learning in English writing: A case from a Chinese university. *Assessing Writing*, 47, 100511. <https://doi.org/10.1016/j.asw.2020.100511>

Zimmerman, B. J., & Schunk, D. H. (2011). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Routledge.