

Guiding Effortful Learning With ChatGPT in University EFL Writing

Miriam T. Black

Toyo Eiwa University

Paul Underwood

Toyo Eiwa University

Reference Data:

Black, M. T., & Underwood, P. (2026). Guiding Effortful Learning With ChatGPT in University EFL Writing. In B. Lacy, M. Swanson, & H. Yoder (Eds.), *LanguageS: Learning, Teaching, Assessing – JALT 50 Years – Challenges and Perspectives*. JALT. <https://doi.org/10.37546/JALTPCP2025-04>

Though concerns have been raised about the misuse of AI such as ChatGPT in academic settings, recent research also emphasizes pedagogical potential when its use is carefully guided. Informed by the concepts of desirable difficulty, productive struggle, cognitive offloading, and the theory of planned behavior, the authors designed exploratory interventions during AY2025 to guide Japanese university EFL students' use of ChatGPT in (a) pre-writing information searches, and (b) AI-supported revision of independently written drafts. Survey data and student assignments highlighted contrasting ways in which AI use was influenced by resource-related factors and task design. The findings suggest how structured, in-class guidance can preserve opportunities for effortful learning while mitigating excessive cognitive offloading to AI. This paper contributes to a theoretical discussion of how AI use can support rather than undermine long-term effortful learning habits in EFL writing.

ChatGPTなどの生成AIについては学術的誤用が懸念されているが、適切な指導の下では教育的な有用性も示されている。本研究は、望ましい困難、生産的葛藤、認知的オフロード、計画的行動理論の概念に基づき、2025年度に日本の大学EFL学生を対象として、①ライティング前の情報探索、②学習者自身が書いた草稿のAI支援による推敲、の2点でChatGPTの活用を指導する探索的な介入を設計した。調査結果と提出された課題の分析から、学習資源の状況や課題設計がAI利用のあり方に異なる影響を与えることが示された。これらの知見から、構造化された授業内指導は、過度な認知的オフロードを抑えつつ、努力を要する学習機会を保持し得ることが示唆された。本稿は、EFLライティングにおける長期的な努力的学習習慣を損なわずにAIを活用する理論的議論に貢献する。

Chatbots are generative AI programs designed to engage users in human-like interaction. Various chatbots are available (Thompson, n.d.), with one of the most prominent being ChatGPT, developed by OpenAI and released in 2022. The already widespread adoption of such tools among students in our Japanese university EFL classes has prompted us to prioritize the critical examination of their pedagogical and ethical implications in our teaching. This inquiry focuses on learners' use of ChatGPT-4o and 5 (free tier), hereinafter referred to as ChatGPT. As ChatGPT-5 was released midway through our inquiry, participants may have used either version. From our preliminary inquiry (Underwood & Black, 2026), several areas of interest for deeper investigation arose. Among them were (a) using ChatGPT to help students in the pre-writing stage more effectively find information and (b) investigating factors that affect the amount and kind of effort students put into the first drafts of their writing assignments, and their use of ChatGPT to provide corrective feedback on these assignments. We created specific interventions for students' use of ChatGPT to investigate these two areas in the 2025 academic year.

Literature Review

Collectively, much of the literature on AI use and learners' written work has foregrounded institutional concerns and ethics (e.g., Ingley & Pack, 2023), learner perceptions of AI feedback (e.g., Escalante et al., 2023; Teng, 2024), and the potential impact of AI on writing and critical thinking skills (e.g., Hading et al., 2024; Hennessy,

JALT2025 Postconference Publication Selected Paper

This paper was chosen as a Selected Paper through consultation with the JALT Publications Board. We express our congratulations to these authors and our appreciation of their well-written paper.

2025; Labadze et al., 2023). More recently, arguments for a theory-driven approach to AI use in language education are also emerging. In the context of university EFL courses, Underwood and Black (2026) drew on insights from cognitive science to suggest that the pedagogical value of AI depends on how it shapes learners' engagement with cognitive effort during the writing process. From this perspective, the authors argued, *desirable difficulty* (Bjork, 1994; Bjork & Bjork, 2011) is especially relevant as it emphasizes learning conditions and tasks that introduce cognitive effort during encoding and retrieval. Such effort has been shown to support more durable learning (e.g., Roediger & Karpicke, 2006) and enhance long-term transfer—the sustained ability to retrieve and apply previously learned knowledge in new and/or more complex contexts over time—and creativity (e.g., Pan & Rickard, 2018). Accordingly, a key concern is whether and how AI-assisted language learning alters the level and nature of cognitive effort learners invest, particularly in relation to opportunities for reflection, revision, and effortful cognitive engagement. Research has also shown that when difficulty exceeds learners' available resources, it may become counterproductive, underscoring the need for instructional conditions that support learners in engaging with difficulty in constructive ways (Kirschner et al., 2006). Here the related concept of *productive struggle* (Hiebert & Grouws, 2007) is relevant. Widely discussed in educational research, this refers to a supported, problem-solving process through which learners engage with difficult tasks. Prior research has suggested that with appropriate scaffolding, guidance, and feedback, productive struggle can develop problem-solving skills and deepen understanding when engaging in cognitively effortful tasks (Baker et al., 2020; Hiebert & Grouws, 2007; Kapur & Bielaczyc, 2012; Warshauer, 2011).

An important corollary of these arguments is the need to consider the specific neural processes engaged by cognitive effort. Recent empirical findings from neuroscience and cognitive science (Kosmyrna et al., 2025), together with theoretical work in learning science (Oakley et al., 2026), suggest that active retrieval and memory consolidation, both of which are facilitated by sustained attention, are neural processes central to long-term learning. When learners engage in *cognitive offloading*, for instance by delegating cognitive tasks to AI rather than engaging neural processes such as remembering (e.g., letting ChatGPT generate text instead of retrieving known grammar and vocabulary) and consolidating (e.g., having ChatGPT create summaries of research articles without later recall or reuse), the reduced level of neural engagement may diminish opportunities for long-term knowledge retention, transfer, and creativity. This work has suggested that using AI in ways that curtail sustained cognitive effort may undermine the conditions necessary for long-term learning.

With the pervasive use of AI among learners, such as in the ways described here, the issue now facing educators is how to maintain the crucial elements of desirable difficulty and productive struggle in the learning process. In this regard, Underwood and Black (2026) argued that the pedagogical value (and ethical use) of AI depends on context-specific conditions that shape learners' capacity and intentions to engage in effortful learning. Emerging from the field of social psychology, the theory of planned behavior (Ajzen, 1985, 2005) posits that intentions to engage in a certain behavior are shaped not only by attitudes, but also by the interplay between perceived social expectations (subjective norms) and resource-related constraints (perceived behavioral control). In the context of university EFL education, the authors observed that although learners may ardently express positive attitudes toward making an effort to write first drafts in English without AI assistance, mounting academic and personal commitments can lead them to rely more heavily on AI as a Japanese-to-English translator, and even author, in order to meet looming deadlines.

Taken together, these perspectives suggest that learners' engagement with AI cannot be understood solely in terms of attitudes or the AI's affordances. Instead, AI engagement must be considered in relation to the cognitive demands of the tasks and the situational constraints that shape learners' behavioral choices. This shift in the field is significant because it moves the discussion from what AI can do for learners in the short term to critical issues of how AI use may shape cognitive engagement, learning processes, and language development over the long term. Importantly, it suggests a lens through which to explore factors that may support and hinder the effective and ethical use of AI in specific contexts of learning.

With these points in mind, the present exploratory study draws on the concepts of desirable difficulty, productive struggle, cognitive offloading, and the theory of planned behavior to frame two inquiries:

- **Inquiry 1 (Black): Pre-Writing** (*finding information with ChatGPT assistance*)
(RQ 1) How do students find trusted, accurate sources of information on an unfamiliar topic, and what difficulties do they encounter?
(RQ 2) In what ways can using ChatGPT complement or improve the effectiveness of their methods?
- **Inquiry 2 (Underwood): Writing** (*independent drafting and AI-supported revision*)
(RQ 3) How much effort do students put into writing their first drafts in English independently of ChatGPT assistance?
(RQ 4) To what extent do students engage with ChatGPT during revision to further enhance their writing and understanding?

(RQ 5) What are students' attitudes toward ChatGPT as an effective study tool, and how are these attitudes mediated by factors such as time constraints and other resources?

Methodology

Data Collection

Prior to conducting our inquiries, and with institutional approval, we obtained informed consent from participants using a research information sheet (in Japanese) that explained the purpose of the research, the data to be collected, and the treatment of personal information. Students indicated their consent to participate by signing and returning an attached form.

Inquiry 1: Pre-Writing (finding information with ChatGPT assistance) explored ChatGPT use in a second-year EFL writing (advanced level) course in the Department of Human Sciences. Students are assigned to classes based on their Computerized Assessment System for English Communication (CASEC) scores. CASEC is an online, computer-adaptive proficiency test that is a widely adopted metric for Japanese university first- and second-year placement (Japan Institute for Educational Measurement, n.d.). This content-based course focused on learning and writing about issues related to social studies, religion, and culture. Sixteen of the 19 students in the class—all Japanese—participated in the inquiry (class CASEC range = 435–780, $M = 521$; CEFR range = upper A2 to lower B1, with an approximate mean of A2+). To address RQ 1 and RQ 2, the following steps were taken (July to November):

1. *Home Research Assignment A*: Students searched for and selected sources of information on their individual topics.
2. *In-Class Survey 1*: Students completed a preliminary survey (in Japanese) via Google Forms. This explored the search methods they had used before the intervention.
3. *In-Class Reference Evaluation Practice*: The teacher and students evaluated sources of information from Home Research Assignment A.
4. *Home Research Assignment B*: Students used the following teacher-created prompt to find sources of information on one set topic: "Tell me the URL links to three high level, trusted, up-to-date sources (in Japanese) of information about the following topic and quote some relevant sentences from each: _____."
5. *In-Class Revised Prompt Practice*: With teacher guidance, students jointly created a list of English-language references for the same set topic in Google Classroom

using a teacher-created revised prompt.

6. *Home Research Assignment C*: Students used the revised prompt to find sources of information on their individual topics, one or more in English.
7. *In-Class Survey 2*: Students completed a follow-up survey (in Japanese) via Google Forms. This explored their attitudes about using the ChatGPT revised prompt, their perceived confidence, and continued difficulties.

Inquiry 2: Writing (independent drafting and AI-supported revision) explored ChatGPT use in a first-year EFL writing (beginner level) course in the Department of Social Sciences. This content-based course focused on learning and writing about international relations and society. All seven students—six Japanese and one Mongolian—participated in the inquiry (CASEC range = 208–273, $M = 238$; CEFR mean = A1). To address RQ 3, RQ 4, and RQ 5, the inquiry comprised these steps, with Steps 2–5 being conducted three times over the course of the first semester (April to July):

1. *Orientation Class*: The students and teacher discussed the potential benefits and disadvantages of AI and the value of using one's own linguistic resources when drafting and revising written work in English.
2. *Independent Draft Writing*: At home, students wrote a 120–150-word report on a content-based topic without any initial ChatGPT (or other AI) support.
3. *Revision with ChatGPT*: At home, students were then instructed to follow a structured protocol with teacher-created prompts (both in Japanese), which sought to provide guided feedback and promote ethical AI use.
4. *In-Class Testing*: Students reproduced reports from memory in a 30–40-minute writing test, followed by reading their report to a partner and Q&A (in Japanese).
5. *In-Class Survey*: Students completed a survey (in Japanese) via Google Forms. Closed and open-response items explored time resources and management, task difficulty, ChatGPT use, and attitudes toward ChatGPT.

Data Analysis

For both inquiries, data from open-response survey items were content analyzed by the researcher in charge of the respective inquiry. In the Results section, data from closed-response items are reported as the raw number of respondents. For Inquiry 1, data on sources of information students used were collected from Home Research Assignments B and C.

Results

Inquiry 1: Pre-Writing

Inquiry 1 explored the use of ChatGPT with 16 EFL advanced-level writing students. It identified students' methods, difficulties, and effects of ChatGPT assistance when finding sources of information for writing assignments.

Search Methods, Perceived Confidence, and Difficulties (RQ 1)

In Survey 1, all students ($N = 16$) reported using search engines and most ($n = 12$) said they had used ChatGPT or other AI when searching for information, with some ($n = 6$) using multiple methods, including library and database searches. Among the choices from a 5-point Likert scale, the majority chose either “somewhat confident” ($n = 6$) or “not so confident” ($n = 7$) in doing this. The main difficulties they indicated were that they could not understand the information in English well enough ($n = 15$) or judge the trustworthiness of online sources in English ($n = 14$). These were followed by it taking too much time ($n = 12$) and not knowing which online sources to trust ($n = 12$).

Intervention Effects, Perceived Confidence, and Continued Difficulties (RQ 2)

Survey 2 was completed by 12 of the 16 participants. After the intervention, the majority of respondents felt either “rather confident” ($n = 4$) or “somewhat confident” ($n = 5$) in finding trusted and accurate sources of information online, with several ($n = 4$) reporting that the intervention specifically was responsible for this. Almost all ($n = 11$) used the revised teacher-created prompt on their own to find sources of information for Home Research Assignment C, and the same number ($n = 11$) explained how using the prompt had made the assignment slightly easier in specific ways (e.g., lessened complexity, streamlined search). However, several ($n = 4$) indicated that they still had difficulty judging the trustworthiness of English-language online sources, and half ($n = 6$) reported having difficulty understanding the information clearly enough in English.

In addition, the majority ($n = 9$) indicated the search was still too time-consuming, with half ($n = 6$) reporting difficulties when using the revised teacher-created prompt, such as not being able to find information on their topic, broken links, or that ChatGPT gave unusable sources of information or ignored APA format. They reported having to search repeatedly, using different methods, or modifying the prompt to find the information they needed, which took time.

For Home Research Assignment B, of the 16 participants, almost half ($n = 7$) used at least one source of information with questionable reliability (i.e., corporate, news, or

fashion magazine site) and few ($n = 3$) used ones in English. However, for Home Research Assignment C, fewer students used questionable sites ($n = 4$), with over half ($n = 11$) including at least one Japanese government site. In addition, all ($N = 16$) provided a source in English, with almost half ($n = 7$) including more than one.

Inquiry 2: Writing

Table 1

Participant Profiles and Observed Patterns in ChatGPT Use

Student	Year	CASEC	Relative Writing Ability	Observed Patterns in ChatGPT Use
s1	2	220	Low	Disciplined; familiarity with academic demands perhaps led to stable effort and less reliance on AI, despite low proficiency.
s2	1	208	Low	Reported limited difficulty but used ChatGPT in basic ways with little evidence of interactive engagement.
s3	1	259	Mid	Demonstrated consistent effort but fatigue and workload constrained sustained interactive engagement with ChatGPT.
s4	1	244	Mid	Persistently time-stressed and fatigued; increasingly used ChatGPT for task completion, indicating overreliance under pressure.
s5	1	273	High	Strongest writer in the group; used ChatGPT strategically for interactive feedback and self- rather than autocorrection.
s6	1	219	Low	Frequently overwhelmed and fatigued; heavy dependence on ChatGPT to manage linguistic demands and complete tasks.

Student	Year	CASEC	Relative Writing Ability	Observed Patterns in ChatGPT Use
s7	1	242	High	Displayed relatively independent and reflective writing behavior; used ChatGPT selectively despite beginner classification.

Note. All students were in the CEFR A1 range, but there were distinct, observable differences in their actual language proficiency levels. Relative Writing Ability refers to the teacher's evaluation of each student's writing ability based on in-class writing exercises and draft-writing tests.

Inquiry 2 explored ChatGPT use with seven EFL beginner-level writing students (see Table 1). It examined their responses to writing first drafts without AI assistance and their subsequent patterns of ChatGPT use during revision.

Effort and Perceived Difficulty (RQ 3)

Across Surveys 1–3, all students ($N = 7$) reported difficulty writing first drafts in English without AI support. This was mainly due to limited vocabulary and grammatical knowledge (control-related factors) rather than difficulty with generating ideas or understanding the topic. Although the proportion of students rating the task as “very hard” decreased slightly over time ($n = 3$ in Survey 1; $n = 2$ in Surveys 2 and 3), grammar and vocabulary remained the most frequently cited sources of difficulty. However, this likely reflected increasing familiarity with the writing process rather than linguistic improvement.

Time, Fatigue, and Writing Effort (RQ 3)

Most students ($n = 6$) reported having sufficient time to work on their assignments outside class, at least in principle. The average first-draft writing time was 1 hour and 37 minutes. However, qualitative responses (translated from Japanese) revealed substantial competing demands (control-related factors), including long commutes, other coursework, test preparation, and part-time employment, with one student (s4) being “quite honestly... overwhelmed.” Sleep duration was also limited for students, with all but one (s1) reporting 5–6 hours per night. Reports of tiredness increased over the semester. Both limited sleep and reported fatigue appeared to constrain students' capacity to sustain independent writing effort when preparing first drafts and perhaps left little time for interactive engagement with the AI for feedback.

ChatGPT Use Patterns (RQ 4)

Student patterns of ChatGPT use varied by writing ability. Stronger writers (s5 and s7) tended to use AI interactively, asking questions or applying feedback to revise their own texts. Weaker writers relied more heavily on autocorrection or translation to complete the task. In addition, there was a noticeable shift over time toward increased reliance on automated correction. The shift coincided with increasing reports of fatigue, limited sleep, and competing academic demands as the semester progressed. Students also expressed positive attitudes toward independent drafting and AI-based feedback, which are discussed below. This pattern suggests that reliance on autocorrection may have reflected constraints on students' cognitive and temporal resources rather than diminished motivation.

Perceived Usefulness of ChatGPT (RQ 5)

Across Surveys 1–3, nearly all students viewed ChatGPT feedback as “very helpful.” However, even for the two strong writers (s5 and s7), qualitative comments suggest that this perceived usefulness often reflected support for task completion and linguistic accuracy rather than sustained engagement with ChatGPT for feedback and extended question-answer interaction. The translated comments of one mid-level writer (s4) reflected this overall impression. The student emphasized that there were limits to how much they could write by themselves and that ChatGPT checked whether their grammar was accurate and helped them when they “couldn't think of the right word.”

Discussion

Inquiry 1 was designed to lessen the cognitive load required of students in an advanced, second-year EFL writing course when searching for information online. This task is complex as EFL learners must carefully choose keywords, create prompts, select the most trustworthy results from hundreds of options, and within each source find relevant information, often using only English.

In the follow-up Survey 2, students reported slightly more confidence in finding trusted and accurate sources of information online. This suggests the task had become less overwhelming for some students. This is also reflected in the improved quality of the sources they found and used (Home Research Assignment C) and their descriptions of the ways that the intervention had made the task easier. In particular, using the prompt narrowed the number of sources they needed to evaluate at one time. In addition, asking ChatGPT to add the page numbers of the quotes in the revised prompt facilitated the

search for relevant information within long documents. As for productive struggle (e.g., Hiebert & Grouws, 2007), the in-class, teacher-guided practice with the revised prompt (Step 5) could also have influenced students' confidence as they received timely answers to questions and hints for solving their problems from the teacher. Furthermore, Home Research Assignment C promoted desirable difficulty (Bjork & Bjork, 2011), evidenced by at least six students who reported problems with ChatGPT's suggestions, yet continued to actively search until they were satisfied with the results.

Despite this, many students still reported that their low language ability hindered them and that assignments were too time-consuming. One factor was the need to select relevant points from English-language sources and paraphrase them. From the viewpoint of productive struggle, breaking this part of the assignment into steps completed in class with teacher guidance over several sessions (e.g., highlighting while reading, making notes of selected information) might reduce the time such assignments require.

Furthermore, the goal of this class was not only to improve students' English writing skills, but also to increase their knowledge of the subject content themes. Though students may find accurate information, they still may not deeply internalize it. Without the deliberate and repeated retrieval practice needed to strengthen relevant neural structures, new information may be less likely to be retained over time. Cognitive offloading, the process of remembering where to find information rather than retrieving the information itself, may therefore reduce opportunities to strengthen relevant neural pathways (Kosmyrna et al., 2025; Oakley et al., 2026). However, with each retrieval, reactivation of the memory pathways can continue for hours afterward (Oakley et al., 2026, p. 579). Therefore, having students write their first drafts independently and use AI only for revision (see Inquiry 2 below), and adding in-class, post-writing retrieval activities should increase the durability of students' learning (Agarwal & Bain, 2019).

Inquiry 2 was designed to encourage desirable difficulty in a beginner, first-year EFL writing course by separating independent draft writing from AI-supported revision. By requiring students to write first drafts without ChatGPT and later reproduce their written work in class from memory (Step 4), the intervention sought to promote effortful retrieval, consolidation of learning, and active language generation. From the perspective of desirable difficulty (Bjork & Bjork, 2011) and productive struggle (Hiebert & Grouws, 2007), these steps were intended to support long-term learning rather than short-term task completion. ChatGPT was positioned not as a solution provider, translator, or author, but as a writing tutor during Step 3 revision. Here, a structured protocol sought to encourage students to identify errors, make their own revisions, and resolve any confusing points of grammar and vocabulary, rather than receiving direct corrections. In

this way, the protocol was intended to maintain appropriate challenges while tailoring support to promote sensemaking and cognitive engagement.

The results indicated that although students consistently experienced independent drafting as difficult, the nature of their engagement with this difficulty varied substantially. Stronger writers in the class were more likely to sustain interactive, reflective use of ChatGPT during revision. In contrast, weaker and more fatigued students increasingly relied on automated correction or translation over the semester. Importantly, for these students this transition was not accompanied by negative attitudes toward effort or learning. Rather, it coincided with reports of limited sleep, competing academic demands, and increasing fatigue.

These findings are consistent with concerns raised in recent cognitive science research regarding cognitive offloading (Kosmyrna et al., 2025; Oakley et al., 2026). When learners are cognitively taxed (i.e., experiencing fatigue and limited attentional, working memory and linguistic resources), AI tools such as ChatGPT may function less as supports for sensemaking and more as mechanisms for reducing immediate cognitive load. This suggests that desirable difficulty, while pedagogically valuable, may be vulnerable in contexts where learners lack the resources to sustain effortful engagement.

While cognitive offloading helps explain *how* students used ChatGPT under pressure, it does not fully account for *why* students with positive attitudes toward desirable difficulty nonetheless struggled to sustain effortful engagement. To more fully account for the gap between students' expressed attitudes and their observed patterns of AI use, the theory of planned behavior (Ajzen, 1985, 2005) provided a useful interpretive framework. Across the three surveys, students expressed positive attitudes toward independent drafting and perceived ChatGPT feedback as highly useful. These findings suggest that motivation and attitudes were not primary barriers. Instead, the decisive factor appears to have been perceived behavioral control. That is, students' capacity to enact their intentions was constrained by low language proficiency, fatigue, time pressure, and limited experience with independent academic work. Under such conditions, even students who valued effortful learning gravitated toward AI use that prioritized task completion over reflective interaction. Without sufficient perceived behavioral control, students' intentions to engage in desirable difficulty may not translate to sustained behavior. This finding thus underscores the importance of considering the situational and structural factors that can shape how learners actually engage with AI tools in practice.

A key pedagogical implication of Inquiry 2 is that ethical and effective AI use cannot be assumed to transfer automatically from protocols (even when carefully crafted) to effective independent study, particularly for first-year EFL students. Unlike the second-

year participants in Inquiry 1, who benefited from more extensive in-class, guided engagement with ChatGPT, the first-year students in Inquiry 2 had limited opportunities to experience AI-supported revision practices under their teacher's supervision. From a theory of planned behavior perspective, for beginner-level university EFL learners such in-class modeling may be critical for strengthening perceived behavioral control and stabilizing their ability to sustain desirable difficulty outside of the classroom.

Conclusion

The capabilities of generative AI are rapidly evolving. Educators are seeking more training, as a crucial part of their jobs is to instill effective learning behaviors in their students. However, it is not clear yet what this training should entail for EFL teachers. In this inquiry, we have drawn on the conceptual frameworks of desirable difficulty and productive struggle, as well as insights from cognitive science and learning theory concerning the importance of durable memory formation and consolidation for creative tasks and long-term learning. However, merely identifying which uses are potentially beneficial does not guarantee that students (and educators) will actually use ChatGPT in these ways. Therefore, we also propose future investigations to examine the specific factors that facilitate or inhibit students' effortful use of AI when learning. In terms of changes to our teaching practices, the findings from this inquiry suggest more guidance in time management, more class time for using ChatGPT in learning tasks, and restructured writing tasks with more in-class written retrieval and review. While not intended for direct generalization to other contexts, this inquiry in a Japanese university EFL setting offers a theory-driven account of how effortful learning in writing can be sustained amid increasingly pervasive AI use among learners.

Bio Data

Miriam T. Black is an associate professor in the Department of Child Education and Care at Toyo Eiwa University, Yokohama. Her areas of interest include examining the role of language in the development of higher thinking processes, especially in preschool-aged children and those with developmental delays. <mblack@toyoeiwa.ac.jp>

Paul Underwood (PhD) is a professor in the Department of International Studies at Toyo Eiwa University, Yokohama and visiting professor at the Graduate School of Language Sciences, Kanda University of International Studies, Tokyo. He teaches CLIL courses (English), ELT methodology, and comparative and international education. <underwood@toyoeiwa.ac.jp>

Authors' Note

ChatGPT was used to refine the authors' original writing, translate student survey responses into English, and translate the writing protocol, surveys, and abstract into Japanese. All content, interpretations, and responsibility for accuracy remain solely with the authors.

The English versions of the Inquiry 1 in-class surveys, the Inquiry 2 ChatGPT writing protocol, and the Inquiry 2 post-writing survey are available from <https://toyoeiwa.repo.nii.ac.jp/records/2000374>.

References

- Agarwal, P. K., & Bain, P. M. (2019). *Powerful teaching: Unleash the science of learning*. Jossey-Bass.
- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action-control: From cognition to behavior* (pp. 11–39). Springer.
- Ajzen, I. (2005). *Attitudes, personality, and behavior* (2nd ed.). Open University Press.
- Baker, K., Jessup, N. A., Jacobs, V. R., Empson, S. B., & Case, J. (2020). Productive struggle in action. *Mathematics Teacher: Learning and Teaching PK-12*, 113(5), 361–374. <https://doi.org/10.5951/MTLT.2019.0060>
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, Article 57. <https://doi.org/10.1186/s41239-023-00425-2>
- Hading, E. F., Rustan, D. R. H. P., & Ruing, F. H. (2024). EFL students' perceptions on the integration of AI in fostering critical thinking skills. *GLENS: Global English Insights Journal*, 2(1), 1–10. <https://doi.org/10.61220/glens.v2i1.466>
- Hennessy, E. (2025). Enhancing L2 writing with ChatGPT for Japanese EFL students. In B. Lacy, M. Swanson, & P. Lege (Eds.), *Moving JALT into the future: Opportunity, diversity, and excellence*. JALT. <https://doi.org/10.37546/JALTPCP2024-04>

- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Information Age Publishing.
- Ingley, S. J., & Pack, A. (2023). Leveraging AI tools to develop the writer rather than the writing. *Trends in Ecology & Evolution*, 38(9), 785–787. <https://doi.org/10.1016/j.tree.2023.05.007>
- Japan Institute for Educational Measurement. (n.d.). CEFR to CASEC sukoa no kankei oyobi meyasu ni tsuite (CEFRとCASECスコアの関係および目安について) [The relationship between CEFR and CASEC scores and approximate benchmarks]. CASEC. <https://casec.evidus.com/materials/cefr/>
- Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, 21(1), 45–83. <https://doi.org/10.1080/10508406.2011.591717>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task*. arXiv. <https://doi.org/10.48550/arXiv.2506.08872>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20, Article 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Oakley, B., Johnston, M., Chen, K.-Z., Jung, E., & Sejnowski, T. (2026). The memory paradox: Why our brains need knowledge in an age of AI. In M. Rangeley & N. Fairfax (Eds.), *The artificial intelligence revolution: Challenges and opportunities* (pp. 573–628). Springer. https://doi.org/10.1007/978-3-032-11814-1_22
- Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, 144(7), 710–756. <https://doi.org/10.1037/bul0000151>
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Teng, M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, 6(3), 36–57. <https://doi.org/10.58304/ijts.20240304>
- Thompson, A. D. (n.d.). *Timeline of AI and language models*. Retrieved January 6, 2024, from <https://lifearchitected.ai/timeline/>
- Underwood, P. R., & Black, M. T. (2026). Exploring ChatGPT use in university EFL courses: Early pedagogical innovations and emerging challenges. *Jinbun Shakai Kagaku Ronshū* (人文・社会科学論集) [Toyo Eiwa Journal of the Humanities and Social Sciences], 43, 37–66. <https://toyoeiwa.repo.nii.ac.jp/records/2000266>
- Warshauer, H. K. (2011). *The role of productive struggle in teaching and learning middle school mathematics* [Doctoral dissertation, University of Texas at Austin]. Texas ScholarWorks. <https://repositories.lib.utexas.edu/items/e824085b-5df9-4ad5-a0e3-e95b79672a2a>