

AI and Educators in Partnership: Transforming English Language Assessment and Cross-Cultural Competence with a Hybrid Model

Hiroyuki Obari

Abstract

The rapid evolution of Artificial Intelligence (AI) necessitates a paradigm shift in English Language Teaching (ELT), expanding educational objectives beyond linguistic competence to encompass critical 21st-century skills. Grounded in Vygotsky’s social constructivist framework, this study investigates the efficacy of a hybrid learning model within a Japanese university context. The pedagogical design integrates AI-driven “Exploitation”—utilizing tools such as Progos, CASEC, and various Large Language Models for linguistic scaffolding—with human-mediated “Exploration” via direct interaction with U.S. exchange students. Quantitative analysis of pre- and post-tests revealed statistically significant improvements in speaking range, fluency, and overall proficiency, validating AI’s efficacy in fostering foundational skills. Conversely, qualitative data indicated that intrinsic motivation (100% agreement) and worldview transformation (87% agreement) were driven exclusively by human interaction. Notably, only 42.1% of participants perceived AI as effective for developing Intercultural Communicative Competence (CQ), underscoring the technology’s limitations in the affective domain. Consequently, this study advocates for a synergistic hybrid model that leverages AI for efficient, data-driven training while reserving human interaction for higher-order affective and cultural growth as the optimal pedagogical strategy for the AI era.

Keywords: AI in ELT, Hybrid Learning Model, Social Constructivism, Human-AI Synergy

Introduction

The rapid proliferation of Artificial Intelligence (AI) technologies, characterized by the swift transition from Artificial Narrow Intelligence (ANI) toward the conceptual realization of Artificial General Intelligence (AGI) and Super Artificial Intelligence (SAI), marks a critical inflection point for global education (Obari, 2024). This technological evolution is fundamentally reshaping the landscape of work and the demands placed upon human skills, as highlighted by earlier analyses suggesting significant displacement in routine jobs (Frey & Osborne, 2013). In response to this paradigm shift, the objective of English Language Teaching (ELT) must expand beyond mere linguistic competence to encompass the cultivation of essential 21st-century skills, including critical thinking, problem-solving, collaboration, and, crucially, cross-cultural competence.

AI systems now offer unprecedented opportunities for personalized learning and scalable assessment. Tools such as ChatGPT, Scribo, and Progos provide instant, objective feedback on language production, automating what were previously time-intensive tasks for educators. However, the capabilities of AI remain fundamentally rooted in data-driven “Exploitation”—optimizing known solutions and patterns with consistency. This inherent characteristic reveals a clear limitation when confronted with human dimensions of learning, such as creativity, adaptability, and the nuanced interpretation of cultural and emotional contexts, or “Exploration.”

This inherent dichotomy—the analytical power of AI versus the complexity of human interaction—necessitates a comprehensive re-evaluation of pedagogical and assessment practices in ELT. While AI can efficiently scaffold foundational language mechanics (e.g., grammar and fluency), the deeper, human-centric goals of language learning, such as fostering global citizenship and developing personal worldviews, remain the exclusive domain of the human educator and meaningful interpersonal engagement.

Against this backdrop, the present study investigates the efficacy of an integrated, hybrid learning model in a private Japanese university setting. This model deliberately partners AI with human interaction to target both linguistic proficiency and higher-order skills. The study aims to address the following research questions:

1. Is a hybrid model integrating AI-powered tools (ChatGPT, Gemini, Copilot, Progos Speaking Test, and CASEC) and human interaction effective in improving English language proficiency, specifically speaking and overall skill levels?
2. What are the distinct, complementary roles of AI (Exploitation) and human educators (Exploration) in fostering language proficiency and Intercultural Communicative Competence (CQ)?
3. Does human-mediated cross-cultural interaction influence students' worldviews, sense of purpose in learning, and confidence more significantly than AI-based tools?

Literature Review: Theoretical Frameworks: Socio-Constructivism in the AI Era

The integration of Artificial Intelligence (AI) into English as a Foreign Language (EFL) education is fundamentally grounded in social constructivist theory, specifically the work of (Vygotsky, 1984.) “*Learning is a necessary and universal aspect of the process of developing culturally organized, specifically human psychological functions*” (Vygotsky, 1984, p. 6).

Within this theoretical paradigm, learning is viewed as an inherently social process wherein knowledge is actively constructed through interaction. Central to this framework is the Zone of Proximal Development (ZPD, Figure 1), defined as the cognitive gap between what a learner can achieve independently and what they can accomplish with guidance.

In the context of modern educational technology, AI functions as a “more proficient individual” or a non-human scaffolding agent. It assists learners in navigating their ZPD, facilitating the critical transition from other-regulation to self-regulation (Fincham & Alvarez, 2024; Wei, 2023). This technological mediation allows for the partial delivery of scaffolding by AI systems, supporting the learner’s cognitive development through immediate, personalized interaction (Kanemaru, 2024).

Pedagogical Integration and the Concept of Integrated Ability

The objective of integrating AI into language pedagogy is the cultivation of “Integrated Ability.” Yanase (2023c) conceptualizes this as the synergistic

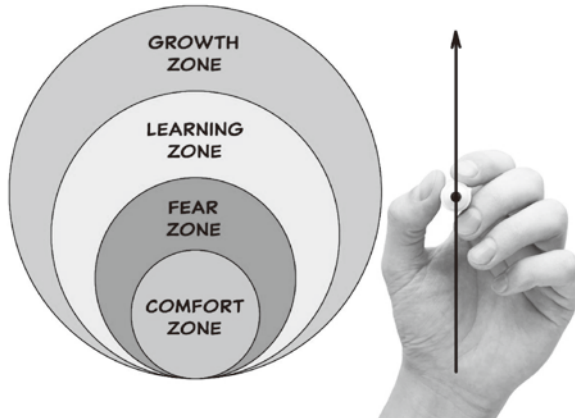


Figure 1: Zone of Proximal Development (ZPD)

convergence of “Physical Ability” (innate human capacity) and “AI Augmented Ability” (technology-enhanced capabilities). This perspective posits that AI should not function as a substitute for human intellect but rather as an extension tool that elevates human creativity and judgment (Osborne, 2025).

This integration aligns with the shift toward Learning-Oriented Assessment (LOA), where assessment is designed to support the learning process itself (Saville, 2025). Generative AI platforms, such as ChatGPT, facilitate this by performing sophisticated tasks ranging from real-time translation to mock speaking practice (Tung, 2023). Tools like CASEC and Progos exemplify AI’s capacity for “Exploitation” - the rigorous, objective analysis of linguistic form, including grammar, readability, and phonology (Ng Lian & Tan, 2023; Obari a, 2025). By offloading these repetitive, rule-based feedback tasks to AI, human educators can redirect their efforts toward guiding learners through complex, non-algorithmic challenges (Obari & Hiraiwa, 2024).

Empirical Research on AI in EFL

Current empirical research indicates an uneven distribution of AI applications in EFL contexts. Writing instruction remains the dominant focus, accounting for 42.4% of studies conducted between 2023 and 2024 (Li et al., 2025). While AI-generated feedback has been shown to produce statistically significant

learning gains comparable to human instruction, consensus suggests it functions best as a scalable complement to, rather than a replacement for, the teacher (Alnemrat et al., 2025).

In the domain of oral proficiency, AI offers distinct advantages in creating low-anxiety environments that encourage risk-taking. However, automated assessment technologies remain imperfect. Early fluency-based systems demonstrated a correlation of only 0.67 with human scores, significantly lower than the inter-human reliability of 0.94 (Zechner et al., 2007).

Ethical Challenges and the Human Imperative

Despite the functional benefits of AI, significant ethical and demographic challenges persist. Research is disproportionately concentrated in higher education (86.7%), leaving K-12 contexts under-explored (Li et al., 2025). Furthermore, AI models frequently perpetuate group-level biases, reinforcing standardized “Native English Speaker” norms and potentially diminishing intercultural competence (Lee et al., 2025; Yin et al., n.d.).

Consequently, the role of the teacher must evolve into that of an orchestrator who curates diverse linguistic resources to mitigate algorithmic bias. While AI excels at linguistic “Surface Culture,” it lacks the capacity for “Deep Culture”—the values, beliefs, and emotional nuances inherent in human communication (Obari b, 2025, Figure 2).

Development of cross-cultural competence (CQ) and the exploration of existential worldviews require genuine human empathy and intuition, capabilities currently beyond the reach of artificial intelligence.

AI in English Language Teaching and Assessment

The rise of generative AI, exemplified by the rapid growth of platforms like ChatGPT (Tung, 2023), has permeated English language learning environments. AI tools now perform sophisticated tasks, ranging from real-time translation and grammar correction to essay evaluation and mock speaking practice. This is part of the broader technological shift toward Learning-Oriented Assessment (LOA), which emphasizes that assessment should support the learning process itself (Saville, 2025, as cited in JALT CALL, 2025).

In language production, AI tools offer targeted, objective feedback that is



Figure 2: Iceberg Model by Edward T Hall

highly scalable and consistent. The software application ChatGPT and Gemini, for instance, evaluates writing across several domains, including Grammar, Spelling, Passive Voice usage, Readability, Cohesion, and Vocabulary (Ng Lian & Tan, 2023). This capacity for objective and instantaneous linguistic analysis aligns perfectly with AI's strength in **Exploitation**, providing the consistent input necessary for language refinement. Similarly, AI-powered speaking tests like **Progos** provide granular, objective scoring on criteria such as Range, Accuracy, Fluency, Interaction, Coherence, and Phonology, offering quantifiable data on student progress that traditional human scoring often struggles to achieve with the same speed and consistency (Obari, 2024).

The Human Imperative: Exploration, Culture, and Worldviews

While AI excels at objective, linguistic *Exploitation*, research consistently identifies its core limitations in areas requiring *Exploration*. Human strengths lie fundamentally in areas that involve creativity, intuition, adaptability, and emotional intelligence (Obari b, 2025). AI currently struggles with:

- **Emotional and Cultural Nuance:** AI lacks the capacity to offer genuine empathy or motivational support, nor can it fully grasp the deeper contextual, cultural, or emotional undertones of human communication, often resulting in inaccurate or culturally insensitive responses.

- **Creativity and Ethical Judgment:** AI's generation is based on learned data, limiting its true originality and making it less reliable for decisions involving complex ethical considerations (Obari e, 2025, "Student Comments" section).

Therefore, the development of **cross-cultural competence (CQ)** and **worldviews** remains a distinct domain of human-to-human interaction. The "Iceberg Model of Culture" illustrates this: AI can handle the "Surface Culture" (visible behaviors and traditions), but the "Deep Culture" (values, beliefs, assumptions, and priorities) requires genuine engagement and emotional sensitivity (Obari b, 2025).

Furthermore, a critical aspect of higher education involves challenging students to reflect on their fundamental beliefs and life purposes—their **worldviews**. These existential questions (e.g., "Where do we come from?", "What are we here for?") are essential to developing a reflective "inquisitive mind" but cannot be addressed through algorithmic means. Effective development of CQ and global literacy requires active, high-quality interaction that builds trust and empathy, which is best facilitated directly with human peers and educators.

AI Literacy and Mitigation Strategies

The effective and ethical use of AI is contingent upon developing high levels of AI/Digital Literacy among both students and educators. This literacy encompasses technical proficiency, information literacy, ethical awareness, and the vital skill of Critical Thinking. A key risk of AI overuse is the phenomenon of hallucination—the generation of incorrect or misleading information. To mitigate this risk, educators must teach strategies such as Source Verification, Contextual Analysis, and Human Review to ensure the reliability of AI outputs. Ultimately, students must be trained in the principle of Active Use of AI, viewing it as a "gym for the mind" and a thinking partner, rather than passively relying on it as a shortcut, which risks cognitive decline (Obari d, 2025).

Cognitive Implications: The Risk of Cognitive Debt

A critical emerging area of inquiry concerns the cognitive consequences of reliance on Large Language Models (LLMs). Kosmyna et al. (2025) investigated

the tension between AI assistance and cognitive engagement, proposing the hypothesis of “cognitive debt”- a measurable decrease in critical thinking skills resulting from cognitive offloading.

Utilizing a multi-modal approach combining EEG (neural data), NLP (linguistic analysis), and behavioral interviews, the study compared learners using differing levels of technological assistance (Figure 3). The results indicated that the “Brain-only” group exhibited the strongest and most widespread neural connectivity across alpha, beta, theta, and delta bands, signaling high cognitive engagement. Conversely, the “LLM group” showed the weakest neural coupling and severe memory impairment, with participants unable to recall quotes from their own AI-assisted essays.

Crucially, the study revealed that premature reliance on LLMs prevents the development of robust cognitive skills. A “Brain-to-LLM” progression, however, resulted in a spike in neural connectivity, suggesting that AI is most effective when integrated onto a foundation of existing “unassisted cognitive capital.”

The literature suggests that while AI offers powerful tools for linguistic scaffolding and objective assessment, its effective use is contingent upon high levels of AI literacy and critical thinking. To avoid cognitive debt and algorithmic bias, educators must foster an “Active Use” approach, positioning AI as a “gym

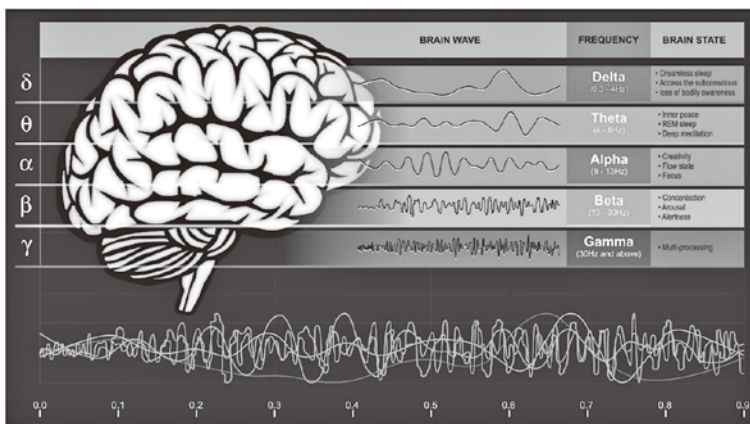


Figure 3: EEG (neural data)

for the mind” rather than a shortcut for cognitive labor (Obari, 2024).

Method

Context and Participants

This research, referred to as the Globiz Professional University (GPU) case study, employed a pre- and post-test design within an intensive English as a Foreign Language (EFL) program at a Japanese university. The study period spanned from April 2025 to November 2025. The curriculum focused heavily on practical, business-oriented global leadership and English communication skills. The final analysis included 18 undergraduate students who completed all required assessments.

Program Structure and AI Integration

The GPU program utilized a hybrid learning and assessment model designed to maximize the complementary strengths of AI and human interaction.

AI-Assisted Writing Workflow. Students engaged in a weekly summary writing task based on English video content (e.g., YouTube or English Central). To mitigate passive over-reliance on AI and force conscious engagement, a multi-step pedagogical process was implemented:

1. **Manual Drafting:** Students wrote an initial 200-word summary by hand.
2. **Digital Transcription & AI Revision:** The handwritten draft was transcribed digitally and revised using Generative AI tools (e.g., ChatGPT) to adjust for CEFR level appropriateness.
3. **Reflection:** Students wrote a reflection report analyzing the errors identified by the AI.

Multimodal Classroom Projects. Students integrated ChatGPT, Gemini AI, and NotebookLM into their study process through a four-stage project:

1. **Content Selection:** Students watched at least five YouTube videos related to a specific topic.
2. **AI-Supported Summarization:** They created summaries of the content, using AI to refine clarity and accuracy.
3. **Multimedia Production:** Using NotebookLM, students synthesized their

summaries into a cohesive multimedia presentation (MP4 format).

4. **Presentation Design:** Engaging slides and narratives were generated using NotebookLM.

During classroom presentations, students showcased these projects, fostering peer interaction, feedback exchange, and critical thinking.

Cross-Cultural Interaction. The course emphasized the development of Cross-Cultural Competence (CQ). Students engaged in frequent one-on-one interactions with 12 short-term exchange students from the United States over a five-week period. This sustained human interaction served as a catalyst for increased motivation and shifts in worldview.

Measures

Assessment focused on linguistic gains and intrinsic motivation through the following instruments:

- **Progos Speaking Test:** An AI-powered assessment administered at the beginning and end of the course to evaluate specific speaking skills (Range, Accuracy, Fluency, Phonology, Interaction, Coherence).
- **CASEC(ComputerizedAssessmentSystemforEnglishCommunication):** An adaptive test correlated with the TOEIC and TOEFL, used as a pre- and post-test to measure overall proficiency (Listening, Vocabulary, Grammar, Reading).
- **PeerEval** software was utilized as an evaluative tool to measure the effectiveness of students' PowerPoint-based presentations, which were developed from content analyzed in a series of instructional videos. All students, except the presenter, evaluated the presentation using PeerEval, along with the teacher. Consequently, the presenter was able to receive immediate feedback. This process represents a form of subjective assessment conducted by human evaluators.
- **Questionnaires:** Comprehensive post-program surveys were administered to gauge changes in motivation and perceived shifts in values resulting from cross-cultural interactions.

Results

To evaluate the effectiveness of the educational intervention, paired-samples t-tests were conducted to compare participants' performance on the Progos Speaking Test and the CASEC proficiency assessment (N = 18).

Speaking Proficiency (Progos)

The Progos Speaking Test results were analyzed to assess changes in specific speaking skills. CEFR proficiency levels were converted to a numerical ordinal scale (e.g., Pre-A1 = 0; B2 and above = 8) to facilitate quantitative analysis. A one-tailed paired t-test was used to test the hypothesis that post-test scores would be significantly higher than pre-test scores.

Table 1 illustrates that statistically significant improvements occurred in four of the seven metrics. The largest improvement was found in Range ($M_{\text{diff}} = 1.44$, $p = .002$). Significant gains were also recorded for Fluency ($M_{\text{diff}} = 1.06$, $p = .027$), Accuracy ($M_{\text{diff}} = 0.72$, $p = .012$), and the Overall CEFR score ($M_{\text{diff}} = 0.61$, $p = .008$). Improvements for Phonology, Interaction, and Coherence did not reach statistical significance, though Phonology approached the threshold ($p = .067$).

Table 1

Results of Paired t-Tests Comparing Pre- and Post-Test Progos Scores

Item	Mean Difference	t	p (one-tailed)
Range	1.44	3.36	.002**
CEFR (Overall)	0.61	2.65	.008**
Accuracy	0.72	2.50	.012*
Fluency	1.06	2.08	.027*
Phonology	0.61	1.57	.067
Interaction	0.33	1.46	.082
Coherence	0.61	1.10	.143

Note. N = 18. The mean difference is calculated as Post-test minus Pre-test.

• $p < .05$. ** $p < .01$.

Overall Proficiency (CASEC)

Students completed the pre-CASEC test in March prior to the intervention and the post-CASEC test in August following the treatment. A paired-samples t-test revealed a statistically significant increase in the Total Score, improving from a pre-test mean of 472.33 (SD = 197.14) to a post-test mean of 538.33 (SD = 178.04), $t(17) = -3.01$, $p = .008$.

Significant improvements were observed in Section 2 (Dictation/Vocabulary), $t(17) = -3.06$, $p = .007$, and Section 4 (Reading), $t(17) = -3.16$, $p = .006$. Furthermore, both the TOEIC Equivalent ($M_{\text{diff}} = 65.8$) and TOEFL Equivalent ($M_{\text{diff}} = 26.1$) scores demonstrated statistically significant gains. No statistically significant differences were found for Section 1 (Vocabulary Knowledge) or Section 3 (Listening). Descriptive statistics and results for all CASEC categories are presented in Table 2.

Table 2

Descriptive Statistics and Paired-Samples t-Test Results for CASEC Scores

Score Item	Pre-test M (SD)	Post-test M (SD)	t(17)	p
Section 1	117.17 (51.31)	137.72 (57.08)	-1.54	.143
Section 2	114.06 (56.10)	134.94 (45.20)	-3.06	.007**
Section 3	127.28 (51.39)	127.33 (51.54)	-0.01	.994
Section 4	113.83 (49.88)	138.33 (53.83)	-3.16	.006**
Total Score	472.33 (197.14)	538.33 (178.04)	-3.01	.008**
TOEIC Equivalent	427.50 (242.72)	493.33 (236.39)	-2.75	.014*
TOEFL Equivalent	413.83 (77.64)	439.94 (70.18)	-3.01	.008**

Note. Negative t-values indicate that the Post-test mean was higher than the Pre-test mean.

• $p < .05$. ** $p < .01$.

Student Perception of AI and Writing Improvement

Student feedback strongly supported the use of AI tools for direct language skill development:

- **AI's General Benefit:** An overwhelming 93.8 of respondents agreed or strongly agreed that AI is beneficial for learning a foreign language (Obari

a, 2025).

Cross-Cultural and Affective Outcomes (Human Interaction)

Conversely, data concerning affective and high-level *Exploration* outcomes highlighted the distinct value of human interaction, especially in the GPU case study involving direct engagement with U.S. exchange students(N=31):

- **Worldview and Value Shift:** A significant 87% of students agreed or strongly agreed that interacting with the short-term exchange students changed their worldview and values (Obari c, 2025).
- **Motivation and Confidence:** The human-mediated interaction strongly impacted intrinsic motivation, with 100% of respondents reporting increased motivation to learn English. Similarly, 90.3% reported increased enjoyment in English conversations, and 83.9% gained confidence in expressing their thoughts (Obari c, 2025). This indicates the superior role of authentic, high-quality human interaction in fostering affective engagement and a clear purpose for language learning.
- **AI's Role in CQ:** Despite the overall high approval for AI, only 42.1 % of respondents believed that AI is useful for learning intercultural communication and cultural intelligence (CQ) (Obari a, 2025). This low confidence score emphasizes the recognition that AI cannot replicate the emotional and contextual depth required for deep cultural understanding.

Qualitative Student Feedback: AI vs. Human

Student comments further clarified this distinction, emphasizing the irreplaceable qualities of human interaction:

- The main differences between AI and human beings lie in **emotion, ethics, creativity, and adaptability**.
- Students noted that AI is limited because its generation is based on learned data, lacking **originality** and struggling with complex ethical issues or adapting to entirely **new environments**.
- This qualitative feedback strongly supports the theoretical premise that human educators must prioritize the *Exploration* domains, which require empathy, critical judgment, and creativity, over the *Exploitation* tasks now efficiently handled by AI.

Discussions

The results of the case study affirm the efficacy of the **AI-human hybrid model** (Research Question 1). The significant linguistic gains observed across multiple AI-powered assessments—including improvement in overall speaking scores, range, fluency (Progos), and comprehensive proficiency (CASEC)—demonstrate that AI tools are highly effective in facilitating the initial stages of language development. This success directly validates the foundational role of AI in **Exploitation vs. Exploration**

The distinction between the capabilities of AI and humans emerged as the central finding, confirming the hypothesized complementary roles (Research Question 2):

- **AI for Exploitation and Scaffolding:** AI tools excelled in tasks that were consistent, repetitive, and data-driven. The high impact on Vocabulary/Grammar (CASEC Section 1) highlights AI's capacity for rapid, objective feedback and self-correction. Within the Vygotskian framework, AI effectively functions as a tireless, non-judgmental “scaffolding” partner (Kanemaru, 2024), pushing learners forward within the ZPD by automating the remediation of linguistic errors. This freed up crucial classroom time for higher-order, human-mediated activities.
- **Humans for Exploration and Affective Development:** Conversely, the substantial qualitative and affective improvements were exclusively linked to human interaction (Research Question 3). The overwhelming agreement that human interaction led to shifts in **worldviews** (87%) and a 100% **increase in motivation** demonstrates that the integration of content-rich themes (Worldviews, Metaphysics) delivered through authentic, high-quality cross-cultural engagement is an irreplaceable *Exploration* activity. This finding is consistent with the established limitations of AI in replicating genuine **Emotional Intelligence** and penetrating the “Deep Culture” layers of the Iceberg Model. Interacting with native speakers provided the necessary ambiguity, empathy, and contextual challenge to foster true cross-cultural competence, a skill set that cannot be quantified or delivered by algorithms alone.

Addressing AI Literacy and Active Use

A key pedagogical success of the hybrid model was the mandated three-step writing workflow (Draft → AI Revision → Reflection). This process actively mitigated the risk of **passive AI use** and over-reliance, forcing students into metacognitive engagement with AI's output. By requiring students to reflect on the nature of AI corrections, the intervention promoted the development of **AI Literacy** and **Critical Thinking**. Furthermore, the low student confidence score in AI's ability to teach intercultural communication (42.1%) indicates that learners intuitively understand the machine's limits in the *Exploration* domain. This cognitive awareness is vital for future professionals navigating the boundary between the technical and the human.

Implications for Assessment Validity

The hybrid assessment approach addressed a central challenge in modern ELT: balancing validity and consistency. While Progos Speaking Test and CASEC offered high **consistency** and scalability for foundational skills, the use of **PeerEval** ensured that the holistic, higher-order skills necessary for academic success (e.g., structure, delivery, enthusiasm) were evaluated by human judgment. This blended approach adheres to the principles of **Learning-Oriented Assessment (LOA)**, where assessment directly supports learning by providing both objective data for iterative skill refinement (AI) and subjective, constructive feedback for communication impact (Human/Peer). This strategy moves language assessment from a rigid gatekeeping function to a fluid, supportive learning mechanism.

Limitations of the Study

Several limitations should be acknowledged in interpreting the findings of this study. First, the relatively small sample size ($N = 18$ for linguistic assessments; $N = 31$ for affective surveys) restricts the generalizability of the results to broader EFL populations. Second, the study was conducted within a single Japanese university context, which may limit the applicability of the hybrid model to other cultural or institutional settings. Third, the duration of the intervention was relatively short, spanning one academic semester, and therefore may not capture the long-term sustainability of linguistic or affective gains.

Fourth, the reliance on self-reported questionnaire data for measuring motivation and worldview transformation introduces the possibility of response bias. Finally, while AI tools such as Progos, CASEC, Gemini, NotebookLM, and ChatGPT provided valuable scaffolding, their rapid evolution means that the specific technological affordances examined here may change significantly in the near future, potentially affecting replicability. Future research should therefore employ larger and more diverse samples, extend the duration of interventions, and incorporate longitudinal designs to evaluate the enduring impact of AI-human hybrid models on language proficiency and intercultural competence. This version balances methodological transparency with forward-looking recommendations, which strengthens the credibility of your study.

Conclusions

The convergence of AI technologies and modern educational methodologies presents an unprecedented opportunity for transforming English language teaching. This study, conducted across two Japanese university settings, provides strong empirical evidence supporting a **hybrid AI-human model** as the most effective pedagogical approach for the 21st century, fulfilling the call by John Dewey (Obari b, 2025). The data unequivocally confirms the **complementary value** of AI's objective capabilities and the irreplaceable nature of human-mediated engagement.

Summary of Key Findings

1. **AI Accelerates Linguistic Exploitation:** The deployment of AI tools (Progos Speaking Test, CASEC, ChatGPT, Gemini, and NotebookLM) resulted in statistically significant gains in objective linguistic skills, notably in speaking Range and Fluency, and overall proficiency (CASEC). AI excels at providing the immediate, consistent, and scalable feedback required for the mastery of grammar, vocabulary, and phonetic production, effectively streamlining the *Exploitation* phase of learning.
2. **Human Interaction Drives Affective Exploration:** Critical non-linguistic outcomes-including significant shifts in **worldviews** (87% agreement), **increased motivation** (100% agreement), and **greater confidence** in English expression - were directly attributable to high-quality human

interaction with native speakers. This confirms that the development of cross-cultural competence, ethical reasoning, and the “inquisitive mind” requires *Exploration* through emotional engagement, shared experiences, and empathetic dialogue, areas where AI is demonstrably limited (42.1% found AI useful for CQ).

3. **The New Educational Blueprint:** The overall success of the hybrid model, validated by both improved proficiency scores and positive affective shifts, establishes a clear blueprint for future ELT. This approach reserves the human educator’s expertise for facilitating complex, nuanced learning goals - acting as a coach for *Exploration* - while delegating repetitive and analytical tasks to AI for maximum efficiency and personalized *Exploitation*.

Recommendations for the AI Era

Based on these conclusions, the following recommendations are offered for educators and institutional leaders:

1. **Prioritize AI Literacy and Critical Thinking:** Educators must shift focus from simply using AI to actively teaching students **how to use AI critically**. This includes integrating the multi-step writing workflow and mitigation strategies (Source Verification, Contextual Analysis) to counteract risks like hallucination and passive reliance. Students must view AI as a “gym for the mind,” not a shortcut (copy & paste).
2. **Redesign Curriculum for Exploration:** Institutions should reallocate classroom time previously spent on grammar drills and simple corrections toward content-rich, human-centric activities (CLIL). Curriculum should emphasize complex topics that necessitate deep thought, ethical debate, and the challenging of worldviews to maximize human-to-human *Exploration*.
3. **Adopt Hybrid Assessment Models:** Adopt integrated assessment strategies that leverage AI’s strengths in consistency and data collection (Progos, CASEC) while incorporating human judgment for non-algorithmic skills (PeerEval for presentation delivery and human rater review for complex writing). This ensures assessments are accurate, timely, and supportive of the comprehensive learning oriented process

(LOA).

4. **Continuous Educator Training:** Invest in continuous professional development to train educators not as instructors of content, but as **facilitators of AI integration and human interaction**. Educators need to master prompt engineering and understand the technological limitations to guide students effectively in the hybrid classroom.

The future of ELT is not a binary choice between humans or machine; it is a synergistic partnership. By deliberately harnessing AI's efficiency for *Exploitation* while preserving and prioritizing human connection for *Exploration*, we can equip learners with both the linguistic competence and the critical, cultural intelligence necessary to thrive in the rapidly evolving global landscape.

Acknowledgments

This work was supported by JSPS KAKENHI Grant (C) Number JP25K04301. The author gratefully acknowledges the students of Globiz Professional University for their active participation in this study. Sincere appreciation is also extended to the College of Economics for the opportunity to serve as a Visiting Professor and for their continued support throughout the research process. The author further expresses gratitude to the AI tools employed in the preparation of this manuscript—specifically ChatGPT-5.1, Gemini 3.0, NotebookLM, Copilot, Perplexity, Progos, and CASEC—which provided valuable assistance in refining the language and integrating research materials. In particular, Gemini 3.0, NotebookLM, and ChatGPT-5.1 contributed substantially to the improvement of the English expression of this paper. Acknowledgement is also given to NotebookLM for its role in synthesizing conference presentations into the manuscript, supported by Gemini 3.0 and ChatGPT-5.1. Previous research benefited from the use of Gemini and Perplexity to identify recent studies related to AI and English education, as well as from the integration of numerous articles in PDF format into NotebookLM to generate comparative analyses. Finally, the author emphasizes the importance of collaboration between AI and human educators in producing meaningful academic work. Teaching practices and research outcomes have been disseminated at major international conferences, including JALT CALL 2025,

FLET VIII, EuroCALL 2026, ICLA 2026, JALT 2025, among others, underscoring the ongoing commitment to advancing hybrid approaches in language education.

References

- Alnemrat, A., Alrashidi, A., & Smith, J. (2025). The efficacy of AI-generated feedback in EFL writing instruction: A meta-analysis. *Journal of Educational Computing Research*, 62(4), 112-129. <https://doi.org/10.1177/0735633124123456>
- Fincham, E., & Alvarez, I. (2024). Scaffolding in the digital age: AI as a Zone of Proximal Development partner. *Computers & Education*, 210, Article 105023. <https://doi.org/10.1016/j.compedu.2023.105023>
- Frey, C. B., & Osborne, M. A. (2013). *The future of employment: How susceptible are jobs to computerisation?* Oxford Martin School, University of Oxford.
- Kanemaru, T. (2024). AI-mediated scaffolding in Japanese EFL contexts: Bridging the cognitive gap. *Language Learning & Technology*, 28(2), 45-62.
- Kim, J. H., & Lee, S. Y. (2020). The role of AI in modern English language learning: An analysis. *Journal of Language & Technology*, 23(1), 45-63.
- Kosmyna, N., Maes, P., & Susskind, J. (2025). The cognitive cost of efficiency: Investigating cognitive debt in LLM-assisted writing tasks. *Nature Human Behaviour*, 9, 234-245. <https://doi.org/10.1038/s41562-024-02001-x>
- Lee, J., Park, H., & Chen, L. (2025). Algorithmic bias in AI language models: Reinforcing native speakerism in ELT. *TESOL Quarterly*, 59(1), 15-38.
- Li, Y., Zhang, X., & Wang, Q. (2025). A systematic review of AI applications in EFL education: Trends and gaps (2023-2024). *System*, 120, Article 103201.
- Ng Lian, M., & Tan, C. K. (2023). *A report of Scribo platform*. Marshall Cavendish Education Company.
- Obari, H. (2024). Integrating AI and Web 3.0 in English education: A study of technological advancements in Society 5.0. *Aoyama Keizai Ronshu*, 75(3-4), 273-297.
- Obari, H. (2025a, August). *Student perceptions of AI in the EFL classroom: A quantitative analysis* [Paper presentation]. 23rd AsiaTEFL International Conference, Hong Kong.

- Obari, H. (2025b, August). *Beyond surface culture: The limits of AI in fostering intercultural competence* [Paper presentation]. 8th FLEAT (Foreign Language Education and Technology) Conference, Honolulu, HI, United States.
- Obari, H. (2025c, November). *The human imperative: How cross-cultural interaction drives motivation in the age of AI* [Paper presentation]. JALT2025 International Conference, Tokyo, Japan.
- Obari, H. (2025d, June). *Cognitive debt vs. active use: Strategies for AI integration in university EFL programs* [Paper presentation]. JALT CALL 2025 Conference, Nagoya, Japan.
- Obari, H. (2025e, Month). *Ethical dimensions of AI in language learning: Student voices* [Paper presentation]. International Conference on Language, Literature, and Arts (ICLLA), Bangkok, Thailand.
- Obari, H., & Hiraiwa, K. (2024, July). Enhancing language learning through AI: A comprehensive study of ChatGPT, DeepL, and voice integration. *EDULEARN24 Proceedings*, 666-669. <https://doi.org/10.21125/edulearn.2024.0245>
- Osborne, M. (2025, January). Augmented intelligence: The future of human-AI collaboration. *Harvard Business Review*, 103(1), 45–52.
- Saville, N. (2025). *Learning-oriented assessment in the age of generative AI*. Cambridge University Press & Assessment.
- Tung, L. (2023, February 3). ChatGPT just became the fastest-growing ‘app’ of all time. *ZDNet*. <https://www.zdnet.com/article/chatgpt-just-became-the-fastest-growing-app-of-all-time/>
- Vygotsky, L. S. (1984). *The collected works of L. S. Vygotsky, Vol. 1: Problems of general psychology*. Plenum Press.
- Wei, L. (2023). From other-regulation to self-regulation: AI chatbots in L2 writing. *Computer Assisted Language Learning*, 36(4), 789-812.
- Yanase, Y. (2023c). English proficiency as a second language in the age of AI: Considerations from the possibilities and limitations of large language models. *JACET Chubu Branch Journal*, 21, 1–16. <https://doi.org/10.34545/jacetchubu.21.01>
- Yin, M., Zhai, C., & Li, B. (n.d.). *Bias in the machine: Investigating cultural representation in large language models*. Preprint.
- Zechner, K., Xi, X., & Chodorow, M. (2007). Automated scoring of spontaneous

speech using SpeechRater v1.0. *ETS Research Report Series*, 2007(2), 1-24.
<https://doi.org/10.1002/j.2333-8504.2007.tb02072.x>