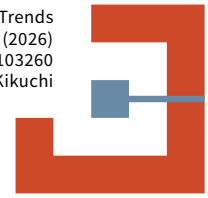
Trends

JALTCALL Trends
ISSN 2982-3749, 2(1), 103260 (2026)
https://doi.org/10.29140/jct.v2n1.103260



Castledown

Developing an AI-use checklist for formative assessment in EMI

Hisayo Kikuchi
Aoyama Gakuin University, Japan
hisayod@gmail.com

Generative artificial intelligence (AI) is reshaping higher education, prompting new concerns about authorship, agency, assessment, and broader pedagogical challenges. In English medium instruction (EMI) contexts, instructors often lack visibility into how students engage with AI during learning tasks. This study introduces an AI-use checklist as a formative assessment tool to document students' AI engagement in an undergraduate EMI course in Japan. Guided by the TPACK framework and Bloom's revised taxonomy, the checklist included ten items covering contextual factors and learning-related dimensions. Findings showed that the checklist effectively distinguished between surface-level patterns and learning-related perceptions. Satisfaction and goal alignment, rather than frequency of use, were most consistently associated with perceived improvements. Qualitative responses also revealed evolving views of AI, with some students describing it as a "group member," highlighting both its collaborative potential and risks to authorship. The study demonstrates that the AI-use checklist offers a practical approach to formative assessment, enabling educators to make AI-use visible and ensure that integration supports learner agency and meaningful engagement in EMI contexts.

Keywords: English medium instruction (EMI), generative AI, formative assessment, TPACK, Bloom's taxonomy

Introduction

The growing presence of AI tools in higher education provides new forms of support for tasks, such as writing, translation, and idea generation, but it also brings concerns about overreliance, ethics, and pedagogical risks (e.g., Zaim et al., 2025). These challenges are particularly acute in English medium instruction (EMI), which refers to the use of English to teach academic subjects in settings where English is not the majority language (Macaro, 2018). Within EMI, language is often treated pragmatically, as a neutral tool not expected to shape disciplinary content (Airey, 2016).

As a result, assessment in EMI has long been described as "problematic" (Dearden,

2014, p. 26), as teachers must balance evaluating subject knowledge with students' uneven English proficiency. The introduction of AI to EMI further complicates these difficulties by raising new concerns, such as academic integrity, authorship, and the evaluation of proficiency (Bannister et al., 2023). In Japan, instructors are often content specialists rather than language teachers; students may have limited English proficiency; and academic expectations are equivalent to those in native-language courses (Kikuchi, 2024). In such contexts, reliance on AI may obscure whether learners are genuinely engaging meaningfully with content and meeting disciplinary goals.

To address these challenges, this study introduces an AI-use checklist as a formative assessment tool to make students' AI engagement visible during a project. Informed by Bloom's revised taxonomy (Anderson et al., 2001) and the Technological Pedagogical Content Knowledge (TPACK) model (Koehler & Mishra, 2009), the checklist captures both contextual and learning-related aspects of AI use. The aim is to assess how effectively the AI-use checklist captures students' engagement with AI and whether it provides meaningful insights for formative assessment.

Method

The study was conducted with 78 students midway through a semester-long EMI semiotics course in Japan. The cohort included 49 domestic and 29 exchange students, of whom 93% were from non-anglophone countries (e.g., Korea, Vietnam, and China).

Working in small groups, students produced a short video as a course project. An explicit AI-use policy permitted AI for brainstorming, drafting, translation, and language support, but prohibited automated image or video generation. The rationale for restricting AI-use was threefold: AI cannot replicate authorial intention (Westberg & Kvåle, 2024), semiotic interpretation requires cultural context and symbolic agency that AI cannot provide (Ghazvineh, 2024), and creative learning involves intentional design rather than automated recombination (Dondero et al., 2025). The checklist served as a formative assessment tool, offering instructors insight into how students engaged with AI and indicating where guidance could support higher-order learning.

The checklist contained ten items guided by Bloom's revised taxonomy (Anderson et al., 2001) and the TPACK framework (Mishra & Koehler, 2006; Figure 1).

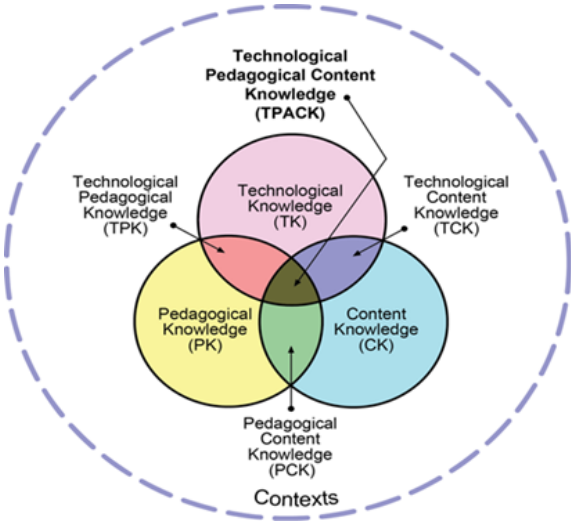


Figure 1. The TPACK model (reproduced by permission of the publisher, © 2012 by tpack.org)

Bloom's taxonomy distinguishes lower-order thinking skills (remembering, understanding, applying) from higher-order skills (analyzing, evaluating, creating). This study adopted the position that AI is best suited to support lower-order skills, while higher-order thinking skills should remain the responsibility of learners to foster cognitive development and agency. The checklist, therefore, documents both surface-level patterns of use and students' perceptions of AI's role in achieving learning outcomes. Four contextual items (Q1–4) captured background factors: frequency, tool choice, satisfaction, and motivation. Six learning-related items (Q5–10) asked about stages of use, perceived impact on work quality, alignment with project goals, collaboration, and support for understanding and expressing content. Table 1 summarizes the items with their corresponding TPACK categories and Bloom's dimensions.

Table 1. AI-use checklist items and their alignment with TPACK and Bloom's taxonomy

Q no	Checklist item	TPACK	Bloom's level
1	How often did you use AI tools?	TK	Contextual
2	What AI tools did you use?	TK	Contextual
3	How satisfied were you with the AI tools you used?	TK	Contextual
4	Why did you choose to use AI in your project?	TPK	Contextual
5	In which parts of the video production process did you use AI?	TPK	Apply/Analyze
6	Do you think AI helped improve the quality of your work?	TPK	Evaluate
7	Did the use of AI match the goals of the group video project?	PCK	Evaluate/Create
8	Was it easy to use AI together with your group?	TPK	Apply/Analyze
9	Did using AI help you better understand the concept of semiotics?	TCK	Understand
10	Do you think AI helped you better express your ideas about the global issue?	TCK	Apply/Create

Questions 1, 3, 6, 7, 9, and 10 were Likert-type items, measured on a five-point scale

ranging from 1 (strongly disagree/very low) to 5 (strongly agree/very high). Survey data were analyzed in SPSS using correlation tests to examine relationships among frequency, satisfaction, work quality, goal alignment, collaboration, and content understanding. The analysis focused on whether outcomes were more strongly linked to the quantity of AI use or to the quality and purpose of its integration.

JALTCALL Trends
Vol. 2 No. 1
2025 JALTCALL Conference Papers Special Issue

Results

Table 2 shows the distribution of responses to Likert-type items.

Table 2. Distribution of responses to Likert-type items on AI use (percentages) (*n*=78)

Q no	Low: 1–2	Neutral: 3	High: 4–5
1	44.9	23.1	32.1
3	6.6	27.6	65.8
6	3.9	18.4	77.7
7	14.9	53.1	59.0
9	14.5	53.5	50.0
10	12.0	33.3	54.3

The majority of students reported positive perceptions of AI use. However, survey analysis showed that the frequency of AI use was not a strong predictor of learning outcomes. Weak to moderate correlations emerged, but they were inconsistent. For example, frequency was modestly linked to improved work quality ($r = .20$, $p = .082$), ease of group use ($r = .24$, $p = .037$), and semiotics learning ($r = .35$, $p = .002$). These results suggest that using AI more often did not necessarily translate into deeper understanding or stronger collaboration.

In contrast, student satisfaction with AI-use was strongly and significantly correlated with multiple outcomes. Higher satisfaction was associated with perceived work quality ($r = .58$, $p < .001$), goal alignment ($r = .64$, $p < .001$), group collaboration ($r = .51$, $p < .001$), semiotics understanding ($r = .58$, $p < .001$), and ease of expressing ideas ($r = .68$, $p < .001$). These findings indicate that perceptions of usefulness and relevance mattered more than frequency of use.

Qualitative responses further showed evolving views of AI's role. Some students described AI as a "group member", suggesting a shift from seeing AI as a neutral tool to a collaborative partner. At the same time, others found it difficult to engage critically with AI-generated suggestions, noting that outputs often appeared "too quickly" to integrate meaningfully.

Discussion

The findings show that the checklist was effective in making students' engagement with AI visible. By distinguishing between contextual and learning-related items, it captured both surface-level usage patterns and students' perceptions of AI's role in learning. This structure allowed analysis beyond simple frequency counts, revealing

that outcomes were shaped more by satisfaction and goal alignment than by the amount of use.

The integration of TPACK and Bloom's taxonomy in the checklist design proved valuable. TPACK ensured attention to technological, pedagogical, and content aspects, while Bloom's hierarchy highlighted the distinction between lower- and higher-order thinking. Students most often relied on AI for lower-order tasks such as translation and grammar, whereas higher-order processes required active engagement and benefited when AI-use was clearly aligned with project goals.

Qualitative responses reinforced this distinction. Some students described AI as a "group member," reflecting its integration into collaborative work, while others expressed difficulty in critically evaluating AI outputs. These contrasting perceptions point to both the potential of AI to enhance collaboration and the risk of diminishing authorship and agency when reliance becomes excessive.

The checklist is also transferable to other AI-supported or CALL courses, providing a practical means of making students' engagement visible.

Conclusion

This study assessed the usefulness of the AI-use checklist as a formative assessment tool for capturing student engagement with AI in classroom practice. The results indicated that frequency of AI-use was not strongly related to outcomes. In contrast, satisfaction and alignment with project goals consistently supported collaboration, understanding, and perceived quality.

The checklist provides a practical means of making engagement visible and can be applied beyond EMI to other AI-supported learning contexts. Future research should extend its use across diverse settings to refine the design and explore how explicit guidance can promote agency-preserving integration of AI.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 25K04297.

References

Airey, J. (2016). EAP, EMI or CLIL? In K. Hyland & P. Shaw (Eds.), *Routledge handbook of English for academic purposes* (pp. 71–83). Routledge.

Anderson, L., Krathwohl, D., & Airasian, P. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. Longman.

Bannister, P., Alcalde Peñalver, E., & Santamaría Urbieta, A. (2023). A systematic review of generative AI and (English medium instruction) higher education. *Aula Abierta, 52,* 401–409. https://doi.org/10.17811/rifie.52.4.2023.401-409

Dearden, J. (2014). *English as a medium of instruction – a growing global phenomenon.* British Council. https://www.britishcouncil.es/sites/default/files/british_council_english_as_a_medium_of_instruction.pdf

Dondero, M., Aldama, J., & Leone, M. (2025). Aspects of AI semiotics: enunciation, agency, and creativity. *Semiotica, 1–3.* https://doi.org/10.1515/sem-2025-0014

Ghazvineh, A. (2024). An inter-semiotic analysis of ideational meaning in text-prompted AI-generated image. *Language and Semiotic Studies, 10*(1), 17–42. https://doi.org/10.1515/lass-2023-0030

Kikuchi, H. (2024). Generative AI as a tool for enhancing English Medium Instruction. In Y. Choubsaz, P. DíezArcón, A. Gimeno-Sanz, J. Hriňák, X. Liashuk, S. Pokrivčáková & H. Vančová (Eds.), *EuroCALL 2024: CALL for humanity* (pp. 87–97). https://doi.org/10.4995/EuroCALL2024.2024.19027

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60–70. https://citejournal.org/volume-9/issue-1-09/general/what-is-technological-pedagogicalcontent-knowledge/

Koehler, M. J., & Mishra, P. (2012). *What is TPACK?* TPACK.org. http://tpack.org

Macaro, E. (2018). *English medium instruction.* Oxford University Press.

Westberg, G., & Kvåle, G. (2024). The generic uniqueness of AI imagery: A critical approach to Dall-E as semiotic technology. *Discourse & Society, 36*(4), 575–598. https://doi.org/10.1177/09579265241274788

Zaim, M., Arsyad, S., Waluyo, B., Ardi, H., Al Hafizh, M., Zakiyah, M., Syafitri, W., Nusi, A., & Hardiah, M. (2024). AI-powered EFL pedagogy: Integrating generative AI into university teaching preparation through UTAUT and activity theory. *Computers and Education: Artificial Intelligence, 7,* 100335. https://doi.org/10.1016/j.caeai.2024.100335

Author biodata

Hisayo Kikuchi is a professor at the School of Global Studies and Collaboration, Aoyama Gakuin University, Japan. Her research focuses on EMI and the integration of AI in higher education. She teaches courses on SLA and sociolinguistics.

JALTCALL Trends
Vol. 2 No. 1
2025 JALTCALL Conference Papers
Special Issue