Andrews, C. (2026). The perceived boundaries of acceptable AI use in EFL writing: Divergent views from students and teachers. *Accents Asia*, *21* (2), 1-29.

The Perceived Boundaries of Acceptable AI Use in EFL Writing: Divergent Views from Students and Teachers

Christopher Andrews[1]

Westgate Corporation

ABSTRACT

This cross-sectional study compared the views of 280 students and 32 teachers at a Japanese university to understand attitudes towards AI use in first- and second-year English EFL writing tasks. Eleven items rated on a four-point academic acceptability scale (1= Completely acceptable, 4= Completely unacceptable) included direct content generation, editing, brainstorming, and translation. Group differences were analysed using Welch's t-tests and Mann–Whitney U tests, with effect sizes reported. The findings reveal substantial divergence between students and teachers, particularly regarding AI translation and AI-generated or heavily AI-edited text, which teachers judged significantly less acceptable. In contrast, lighter, process-oriented uses of AI showed greater agreement. These differing views as well as high levels of intra-group disagreement, reveal the absence of a consensus surrounding acceptable AI use in EFL writing, which suggests a need for clearer institutional guidelines, improved AI literacy. By highlighting divergent perceptions of acceptable AI use among students and teachers, this study contributes to the ongoing debate around acceptable AI use in EFL writing.

Keywords: Generative AI; Academic integrity; AI literacy; CALL; EFL writing

[1] Christopher Andrews holds a master's degree in applied linguistics and TESOL from the University of Portsmouth and has spent the past decade working in education in Japan. He currently teaches at a university in Kanagawa Prefecture, where his recent research has focused on student motivation and AI use in education. Correspondence should be sent to: pt122069bg@jindai.jp.

INTRODUCTION

AI has reshaped how academic writing is taught and assessed, challenging established beliefs about authorship, learning, and assessment. While earlier AI writing tools often produced formulaic or easily identifiable text, recent advances mean AI is now capable of producing fluent text which closely resembles human-written content. This increased sophistication introduces a form of authorship opacity, in which the boundary between human and machine contribution becomes increasingly difficult to identify. At the same time, users may engage in forms of cognitive offloading when using AI, delegating idea generation and language production to the system. As a result, it is increasingly difficult to determine when AI assistance becomes AI authorship.

This problem is especially acute in EFL writing, as in contrast to many other academic domains, writing tasks are designed not only to communicate meaning but also to develop learners' basic writing skills. Processes such as drafting, revising, and reformulating text are therefore integral to language development. When these processes are partially or fully delegated to AI, questions arise not only about authorship, but also about the extent to which the learning process itself is being bypassed.

Within this evolving landscape, students' responses to AI tools are complex and often mixed. While many learners recognise the potential of AI to support aspects of writing such as idea generation, translation, and revision, attitudes toward the appropriateness of these uses can vary considerably (Hossain et al., 2025). Practices such as AI-assisted rewriting are particularly difficult to categorise, as they may be interpreted either as legitimate scaffolding or as the outsourcing of key aspects of writing. Reflecting these concerns, Rahmat et al. (2026) warn that AI must act as a supplement rather than a substitute for student writing. In the absence of prescriptive norms governing AI use in academic writing, emerging research suggests that students and teachers may hold divergent views on what constitutes acceptable use (Neff, Arciaga, & Burri, 2024a). Such differences can lead to disputes over authorship and originality, with potentially serious academic consequences for students.

In recognition of these challenges, policymakers and educational organisations have increasingly emphasised the need for ethical and responsible AI use. For example, UNESCO (2023) has called for the development of institutional strategies and ethical frameworks to guide the integration of generative AI into teaching, learning, and research. However, the absence of universally accepted standards means that evaluations of appropriate AI use are frequently made at the local level, shaped by institutional culture, disciplinary norms, and individual beliefs. This creates a situation in which both teachers and students must navigate ambiguous scenarios without clear or consistent guidance.

Existing approaches to AI governance have therefore been criticised for their limited capacity to address the rapidly evolving and context-sensitive nature of AI use (Batool, Zowghi, & Bano, 2025). In response, scholars have called for more adaptive and flexible frameworks that evolve alongside technological developments (Reuel & Undheim, 2024). At the same time, AI literacy, defined as the ability to critically, effectively, and ethically engage with AI tools, has emerged as an important area of TESOL research (Tour et al., 2025). Despite these developments, there remains limited research examining how AI use boundaries are understood and negotiated within specific instructional contexts, such as EFL writing. While previous studies have explored general attitudes toward AI, less attention has been paid to how teachers and students evaluate specific forms of AI use across different stages of the writing process.

This is significant, as judgments of appropriate AI use may not be uniform across different types of writing activity. For instance, the use of AI for brainstorming or language support may be perceived differently from its use in generating text. Understanding where perceptions align and diverge is therefore critical for the development of contextually appropriate guidelines and pedagogical practices, helping teachers identify areas where the rationale for acceptable and unacceptable AI use may need to be more clearly communicated. Against this backdrop, the present study investigates the attitudes of English major students and their teachers toward the acceptability of AI use in EFL writing tasks within a university context. By identifying patterns of agreement and divergence, and examining the internal consistency of these judgments, this study aims to provide an empirically grounded account of how acceptable common AI use scenarios are perceived to be. In doing so, it seeks to inform the development of clearer, more robust approaches to AI use in EFL writing instruction and assessment.

The study is guided by the following research questions:

RQ1) What patterns of agreement and disagreement exist between students and teachers regarding the acceptability of different AI uses in EFL writing?

RQ2) How internally consistent are students' and teachers' judgments of acceptability across AI-use scenarios, and which AI-use items produce the most intra-group disagreement?

RQ3) What implications do these patterns have for shaping AI use frameworks, AI literacy curricula, and assessment practices in EFL writing?

LITERATURE REVIEW

Generative AI and the changing nature of EFL writing

Generative AI differs from conventional educational tools in that it can produce fluent, contextually appropriate text and is therefore able to participate directly in the writing process itself. Unlike earlier educational technologies, modern AI tools are now embedded across all stages of the writing process, creating a recursive cycle in which both the production of text and its evaluation are increasingly shaped by AI (Guo & Wang, 2024; Li et al., 2024; Teng, 2024). These systems can assist with brainstorming, paraphrasing, translation, revision, and feedback in ways that make human and machine contributions difficult to separate. This blurring of the line between support and substitution is what makes the boundaries of acceptable AI use in writing so difficult to define.

A growing body of research has identified several potential benefits of AI in education. Studies have found it can provide clear feedback, support self-correction, self-evaluation, and self-regulation, as well as enhance speaking abilities and learner engagement and support language learning through the generation of level-appropriate examples (Alghasab, 2025; Kayali et al., 2023; Malik et al., 2023; Nakahara, 2025; Ngo, 2023; Pavlenko & Syzenko, 2024; Qiao & Zhao, 2023; Song & Song, 2023; Teng, 2026). Additionally, previous research has shown that learners may perceive AI tools as supportive and even motivating (e.g., Tsai, 2019; Teng, 2024; Yang, 2025). Taken together, these studies present AI as a valuable resource not only for language support, but also for promoting learner engagement and autonomy.

At the same time, research has increasingly suggested that these benefits are accompanied by pedagogical trade-offs, with concerns raised about overreliance on AI, reduced authenticity in student output, and possible long-term effects on writing development (Alghasab, 2025; H. S. Long, 2024; Huang & Teng, 2025; Yamaoka, 2024). Kosmyna et al. (2025), for example, reported that users of ChatGPT became increasingly reliant on AI output over time, while other studies have noted that students themselves may worry about AI weakening their writing development (Fakir et al., 2024).

This mix of positive and negative considerations has led researchers to characterise teacher attitudes toward AI as inherently ambivalent. Toncelli and Kostka (2024), for example, describe instructors' perspectives as a "love–hate relationship," in which the perceived benefits of AI are tempered by concerns over academic integrity, student overreliance, and the erosion of critical thinking skills. Such findings indicate that AI cannot be understood only in terms of efficiency or convenience, it must also be considered in relation to the learner's role in the writing process and their own ability to responsibly regulate AI use, particularly as AI-assisted feedback has been shown to positively influence revision behaviour and writing quality (Mekheimer, 2025).

Why EFL writing is a distinct context

These concerns are especially significant in EFL writing because writing in this context is not only a product, but also a learning process. In EFL classrooms, drafting, revising, reformulating, and negotiating meaning are central to language development. When AI takes over idea generation, drafting, translation, or rewriting, it may not simply support performance, but may also bypass the very processes through which learners develop fundamental English language skills. This makes EFL writing a particularly important site for examining the boundary between assistance and authorship.

This distinction has led some researchers to question whether the meaning of writing itself is changing in the age of AI. Yeo (2023), for example, suggests that writing instruction may increasingly involve evaluating, selecting, and justifying AI-generated ideas rather than producing all text independently. From this perspective, AI does not merely supplement writing; but is instead increasingly seen as a legitimate and central tool in the writing process. This possibility makes determinations of appropriate AI use in EFL writing especially complex, because the same action may be seen either as legitimate support by some or as a threat to learning by others.

Ambiguity in Acceptable AI Use, Authorship, and Responsible Use

The ambiguity surrounding AI stems in part from the fact that it does not fit neatly into existing categories such as plagiarism, collaboration, or assistance, which themselves can be interpreted differently across institutional and cultural contexts. Chan (2023) reported that while students were familiar with traditional forms of plagiarism, they viewed many AI-related scenarios more ambiguously, while Tour et al. (2025) noted that clear definitions of acceptable AI use are not yet widely shared. As a result, the issue is not limited to copied words, but extends to authorship, disclosure, and the extent of the learner's own intellectual contribution.

The definitional problem is also complicated by cultural variation in how plagiarism and misconduct are understood, as well as by the absence of universal standards (Gullifer & Tyson, 2013). Hysaj, Dean and Freeman (2025) found that some undergraduates justified AI use because of academic pressure and because they did not view the tasks as relevant to their future lives. Price (2024) reported a shift in students' views of ChatGPT, from seeing it as a "cheating tool" to seeing it as a collaborative resource, provided that its output was critically assessed. These studies suggest that judgments about acceptable AI use are not fixed, but depend on users' beliefs about responsibility, learning, and support.

Concerns about authorship and ownership are central to this debate. Bok and Cho (2023) reported that participants felt detached from their writing after using AI tools, and Kosmyna et al. (2025) found that ChatGPT users reported greater detachment from their writing than those who did not use AI. This sense of detachment matters because it may

weaken engagement with the writing process and reduce opportunities for learning. Accordingly, researchers have increasingly emphasised the importance of transparent AI use, clear institutional guidance, and teacher training (Ngo, 2023b; Kohnke & Zou, 2025; Moorhouse & Kohnke, 2024).

A further issue is that responsible AI use depends not only on technical knowledge, but also on learners' ability to judge when and how AI should be used. Gonsalves (2026) emphasises the role of higher-order thinking, while Teng (2025), Singh, Guan, and Rieh (2025), and Wang (2024) collectively highlight the importance of metacognitive regulation, including reflection, monitoring, and self-directed engagement with AI tools. Taken together, these studies suggest that effective AI use is not simply procedural. It relies on learners' capacity to critically regulate their interaction with AI, rather than passively accept its outputs**.** AI-generated responses are often fluent and contextually appropriate yet may still be inaccurate. As a result, research points to the need for learners to actively evaluate the credibility of AI responses and consider how they use it in their work. This indicates that appropriate AI use is not merely a matter of following rules, but requires ongoing judgement, critical evaluation, and awareness of one's own learning processes.

Frameworks, governance, and the need for context-sensitive norms

Various frameworks have been proposed to guide how students should use AI in education (Hamam, 2025; Edmett et al., 2024). Hubbard (2021) also proposed that existing standards can be adapted flexibly to fit different contexts, a point that is highly relevant in EFL writing. However, most policy frameworks remain broad and general, whereas what is considered appropriate AI use often depends on the immediate instructional context. For this reason, governance approaches that focus only on compliance or risk management may be too abstract to address the actual judgments made by students and teachers in specific classrooms. This is why AI literacy, which Tour et al. (2025) define as the ability to understand and engage with the ethical and social implications of AI use, has become increasingly important in the EFL writing context. However, existing frameworks do not always capture these contextual variations, and there is limited understanding of how teachers and students evaluate the acceptability of different forms of AI use in practice.

RESEARCH GAP

Although previous research has examined students' attitudes towards AI (e.g., Mizumoto et al., 2024; Price, 2024; Huang & Teng, 2025; Yamaoka, 2024), teachers' perspectives have received comparatively less attention. Fewer studies still have examined how both groups judge specific forms of AI assistance across the writing process, especially

where the boundary between support and authorship is unclear. As a result, there is limited understanding of where teacher and student views align, where they diverge, and which types of AI use are viewed most ambiguously.

To address this lack of attention, this study investigates the attitudes of English major students and their teachers towards AI use in EFL writing within a Japanese university context. It examines patterns of agreement and disagreement in perceptions of acceptability and explores the extent to which those views are internally consistent across different AI-use scenarios. In doing so, the study aims to provide a context-sensitive account of how appropriate-use boundaries are currently perceived and to inform the development of more context-sensitive guidance. Although findings are most directly applicable to the foreign instructor led first- and second- year English courses at this institution, the methodology is designed for cross-institutional replication and adaptation. Consequently, this investigation complements ongoing theoretical debates in TESOL by providing an institutional snapshot that may inform broader policy development.

ITEM DEVELOPMENT

The 11 survey items (see Appendix) were developed based on prior research and teachers' reported experiences of student AI use, to reflect common AI use behaviours in the department's writing programme. These items were accompanied by a survey introduction which explained the research aims, restated that participation was anonymous and voluntary, and asked participants to respond based on their own beliefs rather than what they perceived rules to be. The items were reviewed by a colleague for clarity and content validity and once verified were translated into Japanese by the researcher for student use. After back translation and translation accuracy verification with a bilingual Japanese colleague to confirm items accuracy and clarity, the items were then piloted with a small group of students, who did not participate in the later survey, to confirm there were no issues in understanding.

For exploratory purposes, the 11 items were loosely grouped into five categories reflecting common AI-use scenarios in writing.

1. Direct AI content generation with minimal editing (Category 1: Items 1-2)
2. AI content generation with substantial human editing (Category 2: Items 3-4)
3. AI as a supportive tool for research and proofreading (Category 3: Items 5,6, 10)
4. AI for brainstorming and outlining (Category 4: Items 7-9)
5. AI for language translation (Category 5: Item 11)

These groupings are intended as a descriptive and exploratory analytical aid and are not presented as validated subscales or underlying constructs. Cronbach's alpha was calculated for the full 11-item scale and for each category to assess internal consistency.

METHODOLOGY

Teachers of first- and second-year English courses within the English department administered the survey to their classes at the end of the 2025 Spring semester. These courses form the foundation of the English course and are taken by almost all students in the department. Both teachers and students were informed that participation was voluntary. Participating teachers were asked to allocate five minutes of class time to complete the survey online. The survey was administered through Qualtrics and accessed via a QR code using students' phones. A Japanese-language explanation was included at the start of the survey, and students answered the questions in Japanese to ensure full understanding, while teacher respondents completed an English-language version of the survey.

The survey consisted of 11 AI use questions, which were rated on a 4-point Likert scale (1 = Completely acceptable, 2 = Somewhat acceptable, 3 = Somewhat unacceptable, and 4 = Completely unacceptable) to avoid neutral responses. While the removal of a neutral option may compel participants to make a choice, in this exploratory case study it was considered that the benefits outweigh the limitations, as it encourages respondents to express a directional judgment, thereby generating more actionable data for subsequent policy development.

To ensure data quality, any responses completed in under 55 seconds (approximately five seconds per item) or responses that selected the same option for all questions were excluded, as these indicated a high probability that respondents had not read the survey items. Two student responses were removed based on these criteria. Participants completed the survey without supervision and were asked not to discuss their answers until completion to reduce response bias, and all data were collected anonymously to protect the privacy of the participants.

In total, 280 student responses were collected. The responses were from a mixture of first- and second-year students who were learning foundational academic English and were placed in classes based on their CEFR level, ranging from A1 to B2. AI use was broadly discouraged by the department, but discussion of AI use and enforcement of this policy were left to individual teachers' discretion. The majority of participants were aged 18–20, and their experience with and awareness of AI varied.

EFL instructors in the university were also invited to complete the survey, through a meeting announcement and a posted sign with a QR code link to the survey. A total of 32 responses were collected, including three responses from teachers who had left the university

in the previous semester but had taught the surveyed student cohort. All teacher respondents were foreign nationals, which is representative of the department's course staff, and the teacher response rate was 88.89% (32/36).

Survey items were measured using a four-point Likert scale and are therefore ordinal in nature. However, given the ongoing debate regarding whether Likert-type responses should be treated as strictly ordinal or as approximately continuous for analysis (Harpe, 2015), a dual analytical approach was adopted. Welch's t-tests were used as the primary means of comparing group responses, as they provide a robust estimate of mean differences and are appropriate where group sizes and variances may differ. This approach is further supported by Mircioiu and Atkinson (2017), who demonstrate that parametric and non-parametric analyses applied to Likert-type data frequently yield comparable substantive conclusions, supporting their combined use as a means of strengthening analytical robustness. Although parametric tests are generally robust to moderate deviations from normality, their performance can be affected when distributions are highly skewed (Stonehouse & Forrester, 1998). Therefore, Mann–Whitney U tests were conducted as a complementary non-parametric analysis to account for both the ordinal properties of the data and potential skewness in the data. This combined approach allowed patterns to be examined across both mean-based and rank-based analyses, with consensus strengthening confidence in the findings and divergence highlighting the need for a more conservative interpretation.

Effect sizes were reported alongside each test (Hedges' *g* for t-tests; Cliff's delta for Mann–Whitney U) to allow for interpretation of the magnitude of observed differences within each analytical framework. Multiple effect size indices were used to align with both parametric (Welch's *t*-tests) and non-parametric (Mann–Whitney U tests) analyses and to ensure that effect magnitudes could be interpreted consistently across differing statistical assumptions regarding data level (interval vs. ordinal) and distributional form (e.g., symmetry vs. an uneven distribution of responses).

All *Hedges' g* values were calculated as teacher minus student, such that positive values indicate that teachers rated items as more unacceptable than students, while negative values indicate that teachers rated items as more acceptable.

Effect sizes were reported using Cliff's delta (δ) for the Mann–Whitney tests, with values interpreted as negligible *(*$|\delta| < .147$), small (.147–.33), medium (.33–.474), or large (>.474). *Cliff's* δ was also calculated as teacher minus student, such that positive values indicate higher teacher scores and negative values indicate higher student scores. Given the 1–4 scale (1 = acceptable to 4 = unacceptable), higher scores indicate lower perceived acceptability.

Ethics

Permission for both the student and teacher surveys was obtained from the departmental ethics committee. To ensure informed consent from students the purpose of the survey was explained before it was administered by each classes' teacher, and a Japanese-

language explanation was included before the survey items began to ensure all participants understood the purpose of the survey. Students were informed that declining to participate would have no effect on their grades, and that they could withdraw from the survey at any time. The anonymity of both groups was maintained throughout data collection and participants were reassured that no personally identifying information would be collected.

RESULTS

After applying the exclusion criteria (responses <55 s; choosing the same answer for all items), two responses were removed and the final sample comprised 280 students and 32 teachers. Each survey item was rated on a 4-point Likert scale (1 = completely acceptable to 4 = completely unacceptable) and Welch's *t*-tests with Hedges *g*, as well as the Mann-Whitney U test with Cliff's delta were then conducted to determine whether the differences between teacher and student responses were statistically significant. All reported effect sizes follow the same sign convention (teacher minus student), with positive values indicating higher unacceptability ratings by teachers. As shown in Table 1, Welch's *t*-test results were examined alongside effect sizes calculated using Hedges' *g*, which adjusts for unequal sample sizes, to interpret group differences more cautiously and minimise the risk of Type I errors.

Table 1 Welch's t-Test Results

Question	Teacher *Mean* (*SD*)	Student *Mean* (*SD*)	Welch's t	*df*	*p*	Hedges *g*	95% CI
1	3.88 (0.41)	3.60 (0.69)	3.3578	53	0.0015	0.4186	0.0522, 0.7849
2	3.31 (0.85)	3.01 (0.86)	1.8891	38	0.0665	0.3484	-0.0175, 0.7143
3	3.06 (0.93)	2.35 (1.00)	4.0588	39	0.0002	0.7131	0.344, 1.0823
4	2.34 (0.85)	2.43 (1.01)	0.5558	41	0.5814	-0.0902	-0.4551, 0.2747
5	1.41 (0.61)	1.65 (0.88)	2.0004	47	0.0512	-0.2794	-0.6449, 0.0861
6	1.69 (0.88)	2.04 (1.03)	2.0921	41	0.0427	-0.3437	-0.7095, 0.0222
7	1.69 (0.81)	1.66 (0.88)	0.1967	39	0.8451	0.0343	-0.3306, 0.3991
8	2.09 (0.84)	2.33 (1.07)	1.4845	43	0.1450	-0.2282	-0.5928, 0.1378
9	2.94 (0.83)	2.52 (1.01)	2.6473	42	0.0114	0.4217	0.0554, 0.7881

10	1.28 (0.57)	1.88 (0.96)	5.1746	53	0.0001	-0.6447	-1.0131,-0.2763
11	3.69 (0.63)	2.76 (1.00)	7.3581	50	0.0001	0.9571	0.5845, 1.3296

Note. *SD* = standard deviation; *t* = Welch's t-statistic; *df* = approximate degrees of freedom; *p* = *p* value; *g* = Hedges' *g* effect size; *CI* = confidence interval**.** Hedges' *g* was calculated as teacher minus student. Because the scale runs 1 = Completely acceptable to 4 = Completely unacceptable, positive g values indicate teachers rated the scenario as less acceptable (gave higher unacceptability scores) than students. Negative *g* values indicate teachers rated the item as more acceptable than students. Ratings are on a 4-point Likert scale (1 = Completely acceptable, 4 = Completely unacceptable). Higher mean scores indicate lower perceived acceptability.

The largest and most meaningful teacher and student differences emerged on Items 3, 10, and 11 (g = 0.71, -0.64, 0.95. These items showed both small *p*-values and large effect sizes, as shown in Table 2, remained statistically significant after applying a Bonferroni multiple-comparison correction, indicating clear group differences on these items. Item 1 also demonstrated a moderate effect ($g \approx 0.42$) and a robust Welch's *t*-test *p*-value, suggesting a reliable group difference.

Table 2 Bonferroni Correction

Question	Original *p*	Adjusted *p*	Significant (Original)	Significant (Adjusted)
1	0.0015	0.0165	Yes	Yes
2	0.0665	0.7315	No	No
3	0.0002	0.0022	Yes	Yes
4	0.5814	1.0000	No	No
5	0.0512	0.5632	No	No
6	0.0427	0.4697	Yes	No
7	0.8451	1.0000	No	No
8	0.1450	1.0000	No	No
9	0.0114	0.1254	Yes	No
10	0.0001	0.0011	Yes	Yes
11	0.0001	0.0011	Yes	Yes

Note. Statistical significance was determined using $\alpha = .05$. Adjusted *p* values greater than 1.00 were truncated to 1.00.

In contrast, several items produced effect sizes below the threshold for statistical significance. Item 9 showed a moderate effect ($g \approx 0.35$), but its *p*-value did not remain significant after correction, indicating that the difference should be cautiously interpreted. Items with smaller effects ($g < 0.30$), including Items 2, 5, 6, and 8, displayed either marginal or non-significant Welch's *t*-test results, suggesting that any apparent differences are small or may reflect Type I errors. Finally, Items 4 and 7 produced negligible effects ($g < 0.10$) and non-significant results, indicating no meaningful difference between teachers and students.

As shown in Table 3, the Mann–Whitney U tests also indicated several significant differences between teachers and students, with effect sizes ranging from negligible to large.

Table 3 Mann-Whitney U Test Results

Question	N_1 Teachers	N_2 Students	*U*	*Z*	*p*	Cliff's δ	Magnitude
1	32	280	3536.0	–1.952	.051	–0.211	Small
2	32	280	3549.0	–1.925	.055	–0.208	Small
3	32	280	2741.5	–3.595	<.001	–0.388	Medium
4	32	280	4298.5	0.374	.711	–0.041	Negligible
5	32	280	3941.5	1.113	.267	–0.120	Negligible
6	32	280	3648.0	1.720	.085	–0.186	Small
7	32	280	4277.0	–0.419	.675	–0.045	Negligible
8	32	280	3954.5	1.086	.278	–0.117	Negligible
9	32	280	3439.0	–2.153	.031	–0.232	Small
10	32	280	2874.0	3.321	.001	–0.358	Medium
11	32	280	2010.0	–5.108	<.001	–0.551	Large

For Items 1 and 2, group differences approached significance ($p = .051$ and $p = .055$, respectively), but effect sizes were small ($\delta = -0.21$ and -0.20), indicating only slightly higher student acceptability. A clearer pattern appeared for Items 3 and 10, where teachers rated these uses as significantly less acceptable than students ($U = 2741.5$, $p < .001$, $\delta = -0.39$; $U = 2874$, $p < .001$, $\delta = -0.36$). These represent medium-sized differences. Item 9 also showed a statistically significant difference ($U = 3439$, $p = .032$, $\delta = -0.23$), though the effect was small. Items 4, 5, 7, and 8 did not show statistically significant differences ($p = .27$–$.71$) and all had negligible effect sizes ($|\delta| = .04$–$.12$), indicating broadly similar teacher and student views for these uses.

The largest divergence occurred for Item 11, where teachers rated the item far less acceptable than students ($U = 2010$, $p < .001$, $\delta = -0.55$), representing a large effect. Across items, all negative δ values indicate that students tended to judge AI uses as more acceptable than teachers. Medium-to-large differences appeared primarily on items involving more extensive forms of AI-assisted text production, whereas items describing lighter or process-oriented AI support produced small or negligible group differences.

As indicated in Table 4, to evaluate the robustness of these item-level differences, internal consistency was then examined using Cronbach's alpha tests which indicated that the scale was reliable for both groups.

Table 4 Cronbach's Alpha Reliability Estimates for Exploratory AI-Use Categories

Group	Category 1: (Direct AI Content Generation and Minimal Editing)	Category 2: (AI Content Generation with Substantial Human Editing)	Category 3: (AI as a Supportive Tool for Research and Proofreading)	Category 4: (AI for Brainstorming and Outlining)
Teachers	$\alpha = .54$ Poor internal consistency	$\alpha = .71$ Acceptable internal consistency	$\alpha = .12$ Poor internal consistency	$\alpha = .72$ Acceptable internal consistency
Students	$\alpha = .74$ Acceptable internal consistency	$\alpha = 0.66$ Questionable internal consistency	$\alpha = .62$ Questionable internal consistency	$\alpha = .81$ Good internal consistency

Students $\alpha = .88$ indicates good internal consistency, demonstrating student responses were relatively homogenous. In contrast, the teachers $\alpha = .76$ was considered acceptable, with moderate consistency. The lower teacher α may be partly due to the smaller teacher sample ($n = 32$) and/or greater heterogeneity among teachers' backgrounds; therefore, scale and subscale reliability estimates for teachers should be interpreted more cautiously.

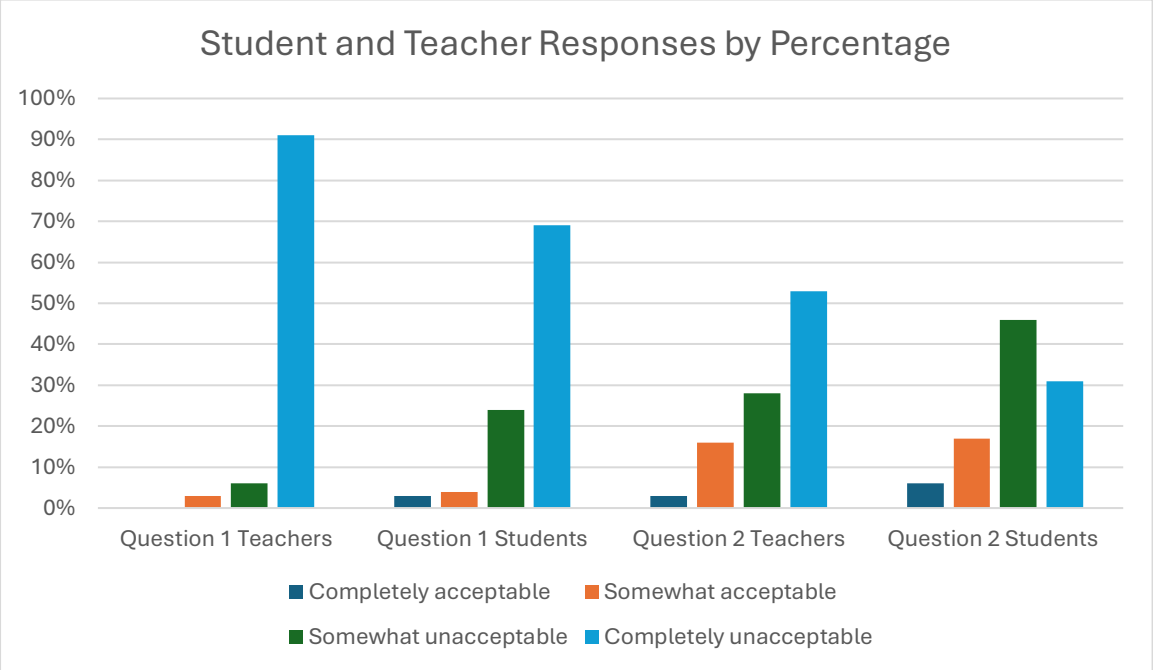
To provide a descriptive overview, the 11 items were tentatively grouped into five AI-use categories. Although the full 11-item scale demonstrated acceptable internal consistency (teachers $\alpha = .76$; students $\alpha = .88$), several category-level reliabilities were suboptimal, particularly for teachers in Category 3 ($\alpha = .12$). Consequently, these exploratory groupings should be regarded solely as descriptive aids rather than as psychometrically validated constructs, and the subsequent category-level analyses warrant cautious interpretation.

Exploratory Category Analysis

1. Direct AI Content Generation with Minimal Editing (Q1–Q2)

As show in Table 5, both groups regarded direct AI generation with little student input as largely unacceptable.

Table 5 Category One Student and Teacher Responses by Percentage

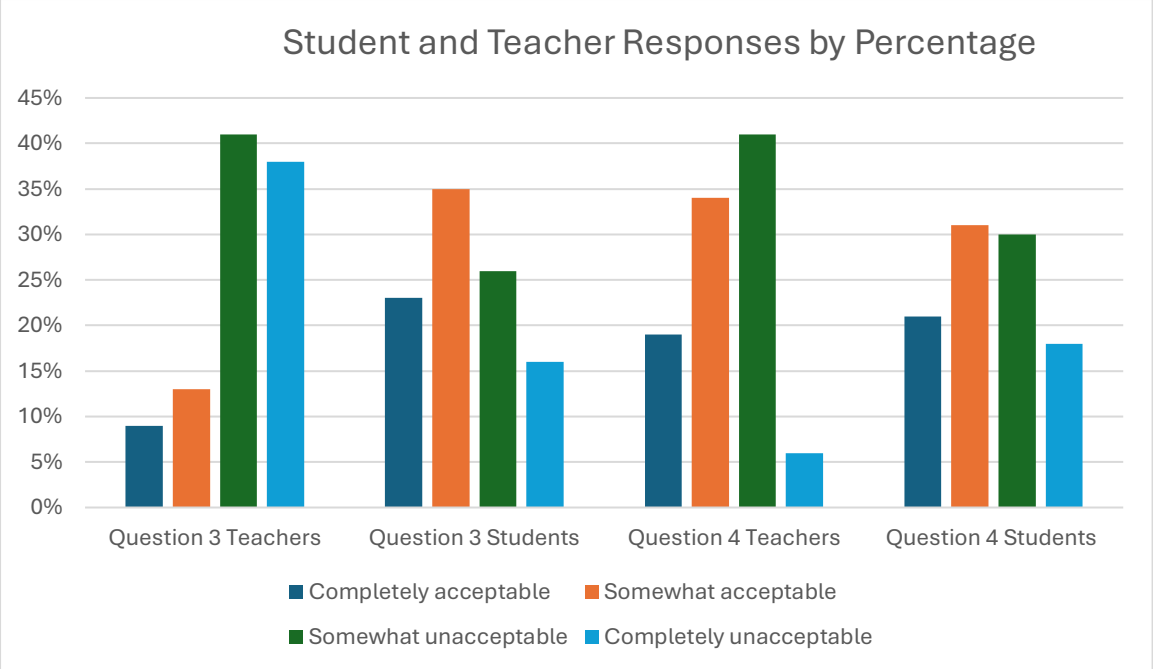


Although teachers rated the direct copy-and-pasting of AI-generated responses into essays (Item 1) and the submission of minimally edited AI-generated content (Item 2) as less acceptable than students did, both groups generally regarded these practices as unacceptable.

2. AI Content Generation with Substantial Human Editing (Q3–Q4)

Responses to this category were more mixed, as demonstrated in Table 6.

Table 6 Category Two Student and Teacher Responses by Percentage

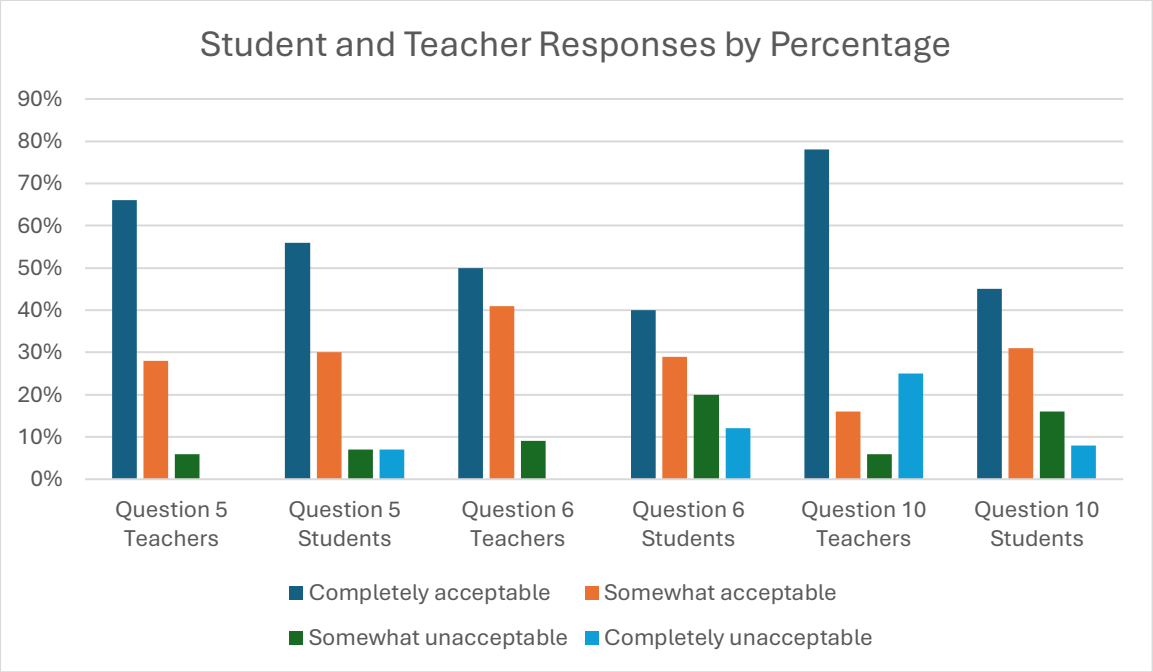


Especially in Item 3, concerning the use of AI-generated answers with substantial formatting and editing, teacher and student views diverged substantially. Students demonstrated a much greater willingness to accept this type of AI use than teachers. Item 4, in which AI was used to rephrase sentences to sound more natural without changing the overall meaning of the content, received mixed responses from both groups, suggesting that judgments became more divided as AI-generated content was edited to a greater extent.

3. AI as a Supportive Tool for Research and Proofreading (Q5–Q6, Q10)

This grouping was the least internally coherent, particularly for teachers, and should therefore be interpreted with caution. Nevertheless, as shown in Table 7, the individual items suggest that lighter, support-oriented uses were generally viewed more positively than direct AI authorship.

Table 7 Category Three Student and Teacher Responses by Percentage

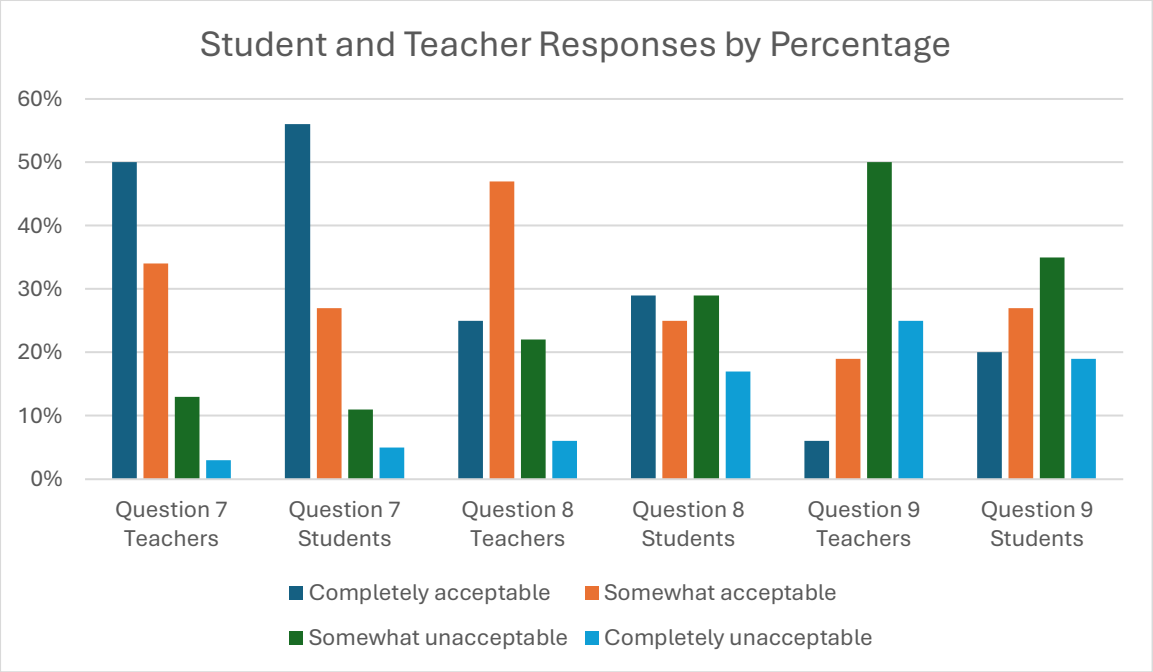


Item 5, the use of AI for spelling and grammar checking, with students making their own corrections, was viewed positively by both groups, although a small number of students indicated that it was completely unacceptable. Item 6, the use of AI to paraphrase or summarise a complex academic article and then use that summary to help write an essay in their own words, was again viewed generally positively, with a minority of students considering this use completely unacceptable. Item 10, the use of AI as a search engine for locating resources, was the only item for which some teachers judged the practice to be completely unacceptable, and notably, a higher proportion of teachers than students expressed this view.

4. AI for Brainstorming and Outlining (Q7–Q9)

As shown in Table 8, responses suggest that early-stage planning uses were generally more acceptable than those involving closer reliance on AI-generated output. This grouping showed relatively stronger internal consistency than some of the others, but should still be treated as exploratory.

Table 8 Category Four Student and Teacher Responses by Percentage

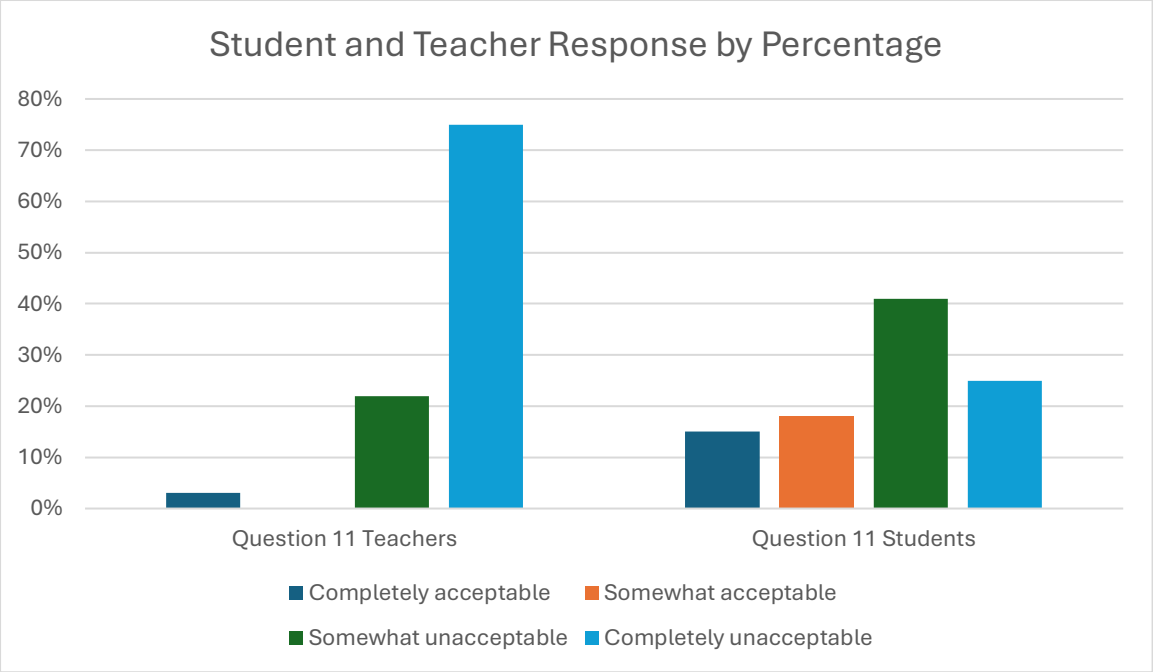


Item 7, the use of AI brainstorming as a starting point for developing students' own ideas, was generally regarded as an acceptable use of AI. Item 8, the selection of the best AI-generated ideas, was more contentious but still generally perceived as acceptable. Meanwhile, the close following of an AI-generated outline in Item 9 was perceived more negatively, especially by the teacher group.

5. AI for Language Translation (Q11)

As indicated in Table 9, Item 11 produced the largest divergence between teachers and students, with teachers viewing wholesale translation much more negatively than students. Because this is a single item rather than a broader subscale, it is best interpreted as one of the clearest points of disagreement in the dataset.

Table 9 Category Five Student and Teacher Response by Percentage



Item 11, in which students brainstormed and wrote their responses in Japanese before translating them into English using AI, was overwhelmingly judged unacceptable by teachers. Students also tended to view this practice as unacceptable, although less decisively, with a minority considering it acceptable. Overall, this item received the strongest negative evaluations, suggesting broad concern about the use of AI for full-text translation rather than more targeted forms of assistance.

DISCUSSION

These findings provide insight into which AI use cases appear to be associated with emerging shared norms and which remain contested. Areas of broad agreement may help inform future policy development at the institution, whereas contested areas may require further clarification to reduce potential confusion regarding acceptable AI use in writing tasks

Perceived Acceptability of AI Use in EFL Writing

A small number of items were areas of strong consensus. The unmodified copying and pasting of AI-generated text (Item 1) was overwhelmingly judged unacceptable by both groups (teacher $M = 3.88$, student $M = 3.60$), indicating that this practice is widely perceived as unacceptable. Item 4, where students used AI to rephrase their own sentences to sound

more natural was contentious but tended to be viewed more positively than negatively by both groups. Meanwhile, item 5 where students use AI to check for errors, but make their own corrections, was viewed as acceptable by both groups (teacher $M = 1.41$, student $M = 1.65$). These results tentatively suggest that items involving more supportive uses of AI received lower unacceptability ratings overall than those involving greater levels of substitution (Items 1–3).

Items 2 and 3, where AI was used to generate text which was edited to varying degrees, indicated increasing ambiguity. Light editing (item 2) appears to introduce some uncertainty, whereas the response to substantial editing (item 3) led to one of the largest discrepancies in the data ($g = 0.71$). Teachers rated Item 3 as largely unacceptable ($M = 3.06$), while students rated it as more acceptable ($M = 2.35$). This pattern, together with the large effect size on Item 11 ($g = 0.96$), indicates that divergent interpretations of the boundary between AI support and authorship exist between the two groups. Similarly, items 6 and 8 demonstrated uncertainty around the use of AI for generating ideas and summarising sources, with moderate-to-high standard deviations in both groups.

A clear divergence in student and teacher perspectives emerges in two items. First, the use of AI to translate Japanese text a student had written into English (item 11) was overwhelmingly rejected by teachers ($M = 3.69$), whereas a significant minority of students felt this was acceptable ($M = 2.76$, $\delta = -0.55$). Second, the close following of an AI generated outline (Item 9), also showed moderate disagreement ($g = 0.42$). These results indicate that divergence is largest on items involving more direct AI contribution to content generation or structure, which may reflect the concerns surrounding diminished authorship reported by (Bok & Cho, 2024; Kosmyna et al., 2025).

Internal Consistency of in Perceptions of Acceptable AI Use

When comparing in-group answers, standard deviation values were relatively high for many items, indicating a lack of internal consensus. This was especially true amongst students, particularly in the case of Items 3 and 11, where significant editing of AI-generated language and translation were surveyed. This lack of stable agreement is consistent with prior research indicating that understandings of AI-related misconduct are often ambiguous and not yet firmly established (Chan, 2023; Gullifer & Tyson, 2013) and demonstrates that norms regarding acceptable AI use in more ambiguous contexts are not firmly established within either group. The implications of such mixed views may include varying expectations across instructors, potentially leading to confusion regarding acceptable practices.

Implications for Policy, AI Literacy, and Assessment in EFL Writing

In line with calls in the literature for the development of institutional strategies and ethical frameworks to guide the use of generative AI in education (UNESCO, 2023; Hamam,

2025), the results reported in this study suggest the need for clearer institutional guidance on AI use, alongside more explicit communication of expectations for both students and teachers. Areas of strong consensus, such as the perceived unacceptability of full AI authorship and the acceptability of limited proofreading support, may provide a useful starting point for such discussions.

Furthermore, these varied responses may also point towards a potential role for AI literacy initiatives in supporting more consistent evaluations of AI use, as called for by Tour et al. (2025). In addition, the findings suggest that certain AI uses which were perceived as appropriate, particularly those involving limited or supportive functions (e.g., using AI for initial drafts followed by documented student revisions or using AI to identify errors followed by self-made corrections), may be permitted for students in writing assignments, with the support of both groups.

Overall, the findings suggest that the greatest differences between teacher and student views emerge in the grey zone where AI use increasingly overlaps with language production. Data from this study suggest that teachers tend to be more restrictive than students when it comes to the submission of substantially revised AI-generated text (Item 3) and using AI for wholesale L1 - L2 translation (Item 11), while a consensus between both groups around the unacceptability of AI authorship (Item 1) and the acceptability of light proofreading help (Item 5) was evident. Taken together, these findings suggest that the boundaries of acceptable AI use in EFL writing vary depending on the type and extent of AI involvement.

Limitations

This study has several limitations that should be acknowledged when interpreting the findings. First, the data was collected from a single university department in Japan, limiting the generalisability of the results to other institutions, disciplines, and cultural contexts. Additionally, the teacher sample was relatively small (n=32) and consisted entirely of English native speakers from various inner-circle nations. Although this teacher sample is representative of the type of teachers encountered by the first- and second-year English course students surveyed in this study, this demographic teacher profile may not reflect the perspectives of Japanese faculty or other foreign national teachers in the same university or other contexts. For instance, one study within Japan has demonstrated that foreign teachers held slightly more positive views of AI use than their Japanese colleagues (Neff et al., 2024b). Given the small and homogenous sample, intra-teacher differences can only be analysed at an exploratory level, so caution is necessary when generalising these findings to other institutional or cultural contexts.

Furthermore, the study did not capture individual variables such as prior experience with AI tools, students' English proficiency level, or teachers' familiarity with generative AI, all of which may contribute to variation within each group. Nor was specific demographic data such as age or gender analysed. Within a different institutional context, different results

could emerge. Moreover, the field of EFL focuses specifically on language creation skills, which may limit the applicability of these findings to other disciplines. Therefore, the results should not be assumed to reflect the views of students and teachers in non-language courses or other courses within the same university.

Additionally, although anonymity was promised, the survey relies on self-reported perceptions of acceptable and unacceptable AI use, which may have been influenced by social desirability bias. Moreover, how closely these self-reported beliefs translate into real world behaviour remains unclear. The internal consistency of the exploratory subcategories was also lower than is necessary to draw reliable conclusions, leading to a reliance on item level analysis. To enable category level analysis, further research should aim to identify more internally consistent subcategories.

Taken together, these limitations underscore the need for future research using multi-institutional samples, mixed-method approaches, and longitudinal designs. While this study contributes to the developing and ongoing global discourse surrounding acceptable AI use, additional research needs to be done to establish broader understanding and unified standards.

Areas for Future Research

The patterns of consensus on fully supportive uses (Items 1 and 5) and divergence on content-generation and translation uses (Items 3 and 11) observed in this study, combined with the mixed views on ambiguous scenarios shown in the data, highlight the need for research that tests the effectiveness of various AI-use policy creation methods. Studies comparing institutional policy development, teacher-led policy creation, and co-creation of AI-use policies with students would clarify which approach produces stronger shared norms and helps to reduce inconsistent assessment practices.

Furthermore, the single-institution sample and the large teacher–student divergence on Item 11 ($\delta = -.55$) indicates value in investigating the extent to which student and teacher attitudes differ across cultures and academic contexts. This study examined Japanese university students and their foreign-national English teachers from inner-circle nations; multi-institutional and cross-cultural replications would establish whether the observed patterns of agreement and divergence on direct AI contribution (Items 3, 9, and 11) hold more broadly and would inform efforts to develop context-sensitive global norms.

Finally, this study did not account for participants' prior experience with AI, levels of English proficiency, or teaching background. Future cross-institutional research incorporating these individual-level variables would offer a more nuanced understanding of variational patterns and assist in the development of effective policies. Such qualitative work could benefit from being designed to minimise social desirability bias by using interviewers with no influence on students' grades, third-person framing, and scenario-based prompts grounded in the specific items used in this survey.

CONCLUSION

Data from this study demonstrate that within the first- and second-year English programmes of the participating university, students and teachers hold divergent views on what constitutes acceptable AI use in EFL writing. Strong consensus appears on Item 1 (unmodified AI-generated text: teacher $M = 3.88$, student $M = 3.60$) and Item 5 (limited proofreading with self-correction: teacher $M = 1.41$, student $M = 1.65$), while significant divergence and high standard deviations characterise many other scenarios (most notably Items 3 and 11, $g = .71$ and $.96$, $\delta = -.39$ and $-.55$). Moreover, while the overall internal consistency within each group was acceptable (students $\alpha = .88$; teachers $\alpha = .76$), suggesting consistent intra-group perceptions, lower category level findings indicate that the perceived acceptability of certain uses of AI remains contentious.

These observed patterns of consensus on clearly supportive uses and divergence on high substitution uses indicate the need for clearly and explicitly delineated AI use rules. Integrating AI literacy into the curriculum could play a useful role in fostering more consistent evaluations of specific scenarios, such as those tested in Items 3 and 11. Finally, as detection of AI use is challenging, and a notable minority of students believe wholesale AI translation is acceptable (Item 11), institutions may wish to consider curriculum changes, such as greater use of in-class assessment, to minimise opportunities for AI misuse.

Looking forward, future research should expand beyond this single institution snapshot to examine whether perceptions of acceptable AI use vary across cultures and institutions. As university education becomes increasingly internationalised, understanding these variations will be essential for developing context sensitive policies, shared norms, preventing inadvertent misuse of AI, and avoiding unfounded accusations of related misconduct.

REFERENCES

Alghasab, M. B. (2025). English as a foreign language (EFL) secondary school students' use of artificial intelligence (AI) tools for developing writing skills: unveiling practices and perceptions. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2505304

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review. *AI Ethics, 5, 3265–3279.* https://doi.org/10.1007/s43681-024-00653-w

Bok, E., & Cho, Y. (2023). Examining Korean EFL college students' experiences and perceptions of using ChatGPT as a writing revision tool. *Journal of English Teaching through Movies and Media, 24(4)*, 15–27. https://doi.org/10.16875/stem.2023.24.4.15

Chan, C. K. Y. (2023). *Is AI changing the rules of academic misconduct? An in-depth look at students' perceptions of "AI-giarism."* arXiv. https://doi.org/10.48550/arXiv.2306.03358

Edmett, A. (2025). GenAI and TESOL. *International Journal of TESOL Studies, 7(2*), 29–36. https://doi.org/10.58304/ijts.250228

Fakir, S. A., Marnaoui, S., & Al Anqodi, H. A. (2024). Written assignments and generative artificial intelligence: Challenges and considerations for English education major students at A'Sharqiyah University, Oman. *Arab World English Journal, 15*(4), 22–38. https://doi.org/10.24093/awej/vol15no4.2

Gonsalves, C. (2026). Generative AI's impact on critical thinking: Revisiting Bloom's taxonomy. *Journal of Marketing Education, 48*(1), 4–19. https://doi.org/10.1177/02734753241305980

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*(7), 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Gullifer, J., & Tyson, G. (2013). Who has read the policy on plagiarism? Unpacking students' understanding of plagiarism. *Studies in Higher Education*, *39*(7), 1202–1218. https://doi.org/10.1080/03075079.2013.777412

Hamam, D. (2025). The AI-resilient educator (AIRE) framework: Implications for educators and TESOL educators. *The Journal of AsiaTEFL, 22*(1), 173–179. https://doi.org/10.18823/asiatefl.2025.22.1.13.173

Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning, 7*(6), 836–850. https://doi.org/10.1016/j.cptl.2015.08.001

Hossain, Z., Çelik, Ö., & Hiniz, G. (2025). Exploring EFL students' AI literacy in academic writing: Insights into familiarity, knowledge and ethical perceptions. *Kuramsal Eğitimbilim, 18*, 157–181. https://doi.org/10.30831/akukeg.1538011

Huang, J., & Teng, M. F. (2025). Peer feedback and ChatGPT-generated feedback on Japanese EFL students' engagement in a foreign language writing context. *Digital Applied Linguistics***,** *2***,** 102469. https://doi.org/10.29140/dal.v2.102469

Hubbard, P. (2021). Revisiting the TESOL Technology Standards for Teachers: Integration and adaptation. *CALICO Journal*, 38(3), 319–337. https://doi.org/10.1558/cj.20068

Hysaj, A., Dean, B. A., & Freeman, M. (2025). Exploring the purposes and uses of generative artificial intelligence tools in academic writing for multicultural students. *Higher Education Research & Development*, *44*(7), 1686–1700. https://doi.org/10.1080/07294360.2025.2488862

Kayali, B., Yavuz, M., Balat, Ş., & Çalışan, M. (2023). Investigation of student experiences with ChatGPT-supported online learning applications in higher education. *Australasian Journal of Educational Technology*, *39*(5), 20–39. https://doi.org/10.14742/ajet.8915

Kohnke, L., & Zou, D. (2025). Artificial intelligence integration in TESOL teacher education: Promoting a critical lens guided by TPACK and SAMR. *TESOL Quarterly, 59*(S3), S267–S278. https://doi.org/10.1002/tesq.3396

Kosmyna, N., Hauptmann, E., Yuan, Y., Situ, J., Liao, X.-H., Beresnitzky, A., 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

Li, J., Huang, J., Wu, W., & Whipple, P. B. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities and Social Sciences Communications, 11*, 1268. https://doi.org/10.1057/s41599-024-03755-2

Long, H. S. (2024). Exploring the use of ChatGPT as a tool for written corrective feedback in an EFL classroom. *The Journal of AsiaTEFL*, *21*(2), 397–412. https://doi.org/10.18823/asiatefl.2024.21.2.8.397

Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki. (2023). Exploring artificial intelligence in academic essays: Higher education students' perspectives. International Journal of Educational Research Open, 5, 100296. https://doi.org/10.1016/j.ijedro.2023.100296

Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. *Discover Education, 4*, 170. https://doi.org/10.1007/s44217-025-00602-7

Mircioiu, C., & Atkinson, J. (2017). A comparison of parametric and non-parametric methods applied to a Likert scale. *Pharmacy*, *5*(2), 26. https://doi.org/10.3390/pharmacy5020026

Mizumoto, A., Yasuda, S., & Tamura, Y. (2024). Identifying ChatGPT-generated texts in EFL students' writing: A comparative analysis of linguistic fingerprints. *Applied Corpus Linguistics, 7*, 100106. https://doi.org/10.1016/j.acorp.2024.100106

Moorhouse, B. L., & Kohnke, L. (2024). The effects of generative AI on initial language teacher education: The perceptions of teacher educators. *System, 122*, 103290. https://doi.org/10.1016/j.system.2024.103290

Nakahara, R. (2025). Is data-driven learning with AI-generated example sentences effective for elementary-level learners? A case study of the synonyms "collect" and "gather." *Asian Journal of English Language Teaching, 34*(1), 5–22. https://doi.org/10.65961/AJELT-2025-1-002

Neff, J., Arciaga, K., & Burri, M. (2024a). EFL students' and teachers' perceptions of the ethical uses of AI tools. *Technology in Language Teaching & Learning*, *6*(3), 1714. https://doi.org/10.29140/tltl.v6n3.1714

Neff, J., Arciaga, K., & Burri, M. (2024b). The impact of EFL teachers' demographic backgrounds on the perceptions of ethical AI use. *PanSIG Journal, 10*(1), 60–66. https://doi.org/10.37546/JALTPanSIGJ10.1-8

Ngo, T. T. A. (2023). The perception by university students of the use of ChatGPT in education. *International Journal of Emerging Technologies in Learning (iJET)*, *18*(17), 4–19. https://doi.org/10.3991/ijet.v18i17.39019

Pavlenko, O., & Syzenko, A. (2024). Using ChatGPT as a learning tool: A study of Ukrainian students' perceptions. *Arab World English Journal*, *1*(1), 252–264. https://doi.org/10.24093/awej/chatgpt.17

Price, G. (2024). Japanese university EFL student insights on the emerging ChatGPT phenomenon. *International Journal of Education, Learning and Development, 12*(5), 83–100. https://doi.org/10.37745/ijeld.2013/vol12n583100

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1255594

Rahmat, N. H., Khairuddin, Z., Aripin, N., & Mokhtar, M. I. (2026). Scaffolding academic writing using AI assistance: A systematic literature review. *International Journal of Research and Innovation in Social Science, 10*(1), 7263–7280. https://doi.org/10.47772/IJRISS.2026.10100560

Reuel, A., & Undheim, T. A. (2024). *Generative AI needs adaptive governance.* arXiv. https://doi.org/10.48550/arXiv.2406.04554

Singh, A., Guan, Z., & Rieh, S. Y. (2025). Enhancing critical thinking in generative AI search with metacognitive prompts. Proceedings of the Association for Information Science and Technology, 62(1), 672–684. https://doi.org/10.1002/pra2.1287

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1260843

Stonehouse, J. M., & Forrester, G. J. (1998). Robustness of the t and U tests under combined assumption violations. *Journal of Applied Statistics*, *25*(1), 63–74. https://doi.org/10.1080/02664769823304

Teng, M. F. (2024). ChatGPT is the companion, not enemies: EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. https://doi.org/10.1016/j.caeai.2024.100270

Teng, M. F. (2025). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education, 60*(1), e12811. https://doi.org/10.1111/ejed.12811

Teng, M. F. (2026). The impact of GenAI feedback on Chinese EFL students' emotional engagement: A mixed effect modelling approach. *Asia-Pacific Education Researcher*. https://doi.org/10.1007/s40299-025-01067-w

Toncelli, R., & Kostka, I. (2024). A love-hate relationship: Exploring faculty attitudes towards GenAI and its integration into teaching. *International Journal of TESOL Studies, 66*(3), 77–94. https://doi.org/10.58304/ijts.20240306

Tour, E., Pegrum, M., & Macdonald, S. (2025). Engaging English language learners in AI literacy practices: a conceptual framework and practical strategies for educators. *English Australia Journal*, *41*(1), 27–46. https://doi.org/10.61504/YBUY9086

Tsai, S.-C. (2019). Using Google Translate in EFL drafts: a preliminary investigation. *Computer Assisted Language Learning, 32*(5-6), 510–526. https://doi.org/10.1080/09588221.2018.1527361

UNESCO. (2023). *Guidance for generative AI in education and research*. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Uğraş, H., Uğraş, M., Papadakis, S., & Kalogiannakis, M. (2025). Innovative early childhood STEM education with ChatGPT: Teacher perspectives. *Technology, Knowledge and Learning, 30*, 809–831. https://doi.org/10.1007/s10758-024-09804-8

Wang, Y. (2024). Cognitive and sociocultural dynamics of self-regulated use of machine translation and generative AI tools in academic EFL writing. *System*, *126*, 103505. https://doi.org/10.1016/j.system.2024.103505

Yamaoka, A. (2024). ChatGPT's motivational effects on Japanese university EFL learners: A qualitative analysis. *TESOL International Journal, 19(2*), 117–132. https://www.tesolunion.org/journal/details/info/6MjQ4dLjhl

Yang, S. (2025). Understanding EFL learners' excessive use of translation software: An extension of the flow theory. *PLOS ONE*, *20*(11), e0335610. https://doi.org/10.1371/journal.pone.0335610

Yeo, M. A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal*, *14*(3). https://doi.org/10.1002/tesj.716

APPENDIX

Teacher Survey

Thank you for taking part in this survey on teacher attitudes toward student use of AI in writing assignments.

You will be presented with a series of scenarios involving student use of AI tools. For each one, please indicate the extent to which you believe the use of AI should be considered acceptable in the absence of specific instructions from the university or course staff. Please respond based on your personal views, not on what you believe current university policies are.

To help maintain the integrity of the survey, please complete it independently and refrain from discussing your responses with colleagues while the survey is ongoing.

1 – The student used AI to generate an answer, then copy and pasted the answer into their essay and submitted it as their own work.

2 – The student used AI to generate an answer, then edited the sentences and formatted a little before submitting

3 – The student used AI to generate an answer, then edited the sentences and formatted significantly before submitting.

4 - The student used AI to rephrase their sentences to make them sound more natural, without changing the overall meaning of the content.

5 - The student used AI to check their grammar and spelling but made their own corrections.

6 - The student used AI to paraphrase or summarize a complex academic article they found difficult to understand, then used that summary to help them write their essay in their own words.

7 - The student used AI to help them brainstorm ideas and used that as a starting point to think of their own ideas.

8 - The student used AI to help them brainstorm ideas and then used the best ones.

9 - The student asked AI to write an outline for their essay and then followed it closely.

10 - The student used AI as a search engine to find resources.

11 - The student did all the brainstorming and writing by themselves in Japanese, then asked the AI to translate their work into English.

Student Survey

このアンケートにご協力いただき、ありがとうございます。本調査は、学生による AI の使用に対する意識を把握することを目的としています。

これから、学生が AI ツールを使用するさまざまな場面が提示されます。それぞれの場面について、大学や教員から特別な指示がない場合に、その AI の使用がどの程度許容されるべきだと思うかをお答えください。ご自身の個人的な考えに基づいてお答えください。現在の大学の方針に基づく必要はありません。

アンケートの信頼性を保つために、必ず一人で回答し、アンケート実施中は他の学生と回答内容について話し合わないようにしてください

これらの質問には正解はありません。これはあくまで皆さんの意見を調査するためのアンケートです。回答は匿名で行われ、最終成績や担当教員の評価には一切影響しません。

1- 学生は AI を使って解答を作成し、それをそのままコピー＆ペーストして、自分の作品として提出した。

2- 学生は AI を使って解答を作成し、文章や形式を少しだけ編集してから提出した。

3- 学生は AI を使って解答を作成し、文章や形式を大幅に編集してから提出した。

4- 学生は AI を使って自分の文章をより自然な表現に言い換えたが、内容の意味自体は変更しなかった

5- 学生は AI を使って文法やスペルのチェックを行い、修正は自分で行った。

6- 学生は理解が難しい学術論文を AI に要約・言い換えさせ、その要約を参考にして自分の言葉でエッセイを書いた。

7- 学生はアイデアを考えるために AI を使い、それを出発点として自分なりのアイデアを考えた

8- 学生は AI にアイデアを出させ、その中から良いものを選んで使用した。

9- 学生は AI にエッセイの構成（アウトライン）を作らせ、それに従って執筆した。

10- 学生は AI を検索エンジンとして使用して資料を探した。

11- 学生は日本語で自分でアイデアを出し、執筆を行った後、その内容を AI に英語に翻訳させた。