SALUS CULTURA
JURNAL PEMBANGUNAN MANUSIA DAN KEBUDAYAAN

ARTICLE

OPEN ACCESS

Governing Classroom AI: Transparency, Integrity; and Learning in TEFL/EAP

Menata Tata Kelola AI di Kelas: Transparansi, Integritas dan Pembelajaran dalam TEFL/EAP

Estiningtyas Sholikhah[1], Lutfiana[2]
[1, 2] Universitas Muhammadiyah Brebes, Indonesia

Corresponding Author
Estiningtyas Sholikhah
estiningtyas@umbs.ac.id

Article History
Received: October 24, 2025
Final Revision: December 15, 2025
Accepted: December 16, 2025
Available Online: December 31, 2025

Abstract

The rapid uptake of generative AI in TEFL and EAP has outpaced course-level governance, particularly with respect to transparent AI use, academic integrity, and data protection. This study aims to develop and evaluate a human-centred AI Classroom Contract as an auditable micro-policy that bridges the gap between guidance and classroom practice. Over eight months, we conducted a design-based study combined with a quasi-experimental comparison across classes with three collection points: baseline, mid-course, and end. Data included per-assignment AI transparency forms, learning-management-system logs on revision iterations, time on task, and punctuality, draft–revision artefacts with writing and speaking rubric scores, integrity audits, and surveys on perceived fairness and transparency. Analyses used multilevel modelling and mediation tests. Findings indicate higher completeness of transparency statements, fewer mis-citation and fabrication incidents, more iterative revision, improved punctuality, and moderate gains in task performance. Perceived fairness and transparency mediated effects on compliance and outcomes. Implementation was feasible under limited connectivity without meaningful privacy breaches. We conclude that the AI Classroom Contract is an effective micro-policy instrument that connects human-centred principles to TEFL and EAP practice, yielding a replicable package of the contract, templates, and a fact-check rubric to support quality and accountability in language learning.

Keywords

Academic integrity; TEFL; EAP; AI Classroom Contract; Learning Analytics

Abstrak

Adopsi kecerdasan buatan generatif pada pembelajaran TEFL dan EAP meningkat pesat, sementara tata kelola di tingkat mata kuliah, khususnya transparansi penggunaan AI, integritas akademik, dan perlindungan data, masih belum seragam. Penelitian ini bertujuan mengembangkan serta mengevaluasi Kontrak Kelas AI berprinsip human-centred sebagai kebijakan mikro yang dapat diaudit untuk menjembatani kesenjangan antara pedoman dan praktik. Studi berlangsung delapan bulan dengan pendekatan berbasis desain yang dipadukan dengan perbandingan kuasi-eksperimental antarkelas pada tiga titik pengumpulan data: awal, pertengahan, dan akhir. Data mencakup formulir transparansi per tugas, log LMS tentang iterasi revisi, waktu belajar, dan ketepatan unggah, artefak draf dan revisi dengan skor rubrik menulis maupun berbicara, audit integritas, serta survei persepsi keadilan dan transparansi. Analisis menggunakan pemodelan multilevel dan uji mediasi. Hasil menunjukkan peningkatan kelengkapan pernyataan transparansi, penurunan insiden mis-citation dan fabrikasi, lebih banyak iterasi revisi, ketepatan waktu yang lebih baik, serta kenaikan moderat pada kinerja tugas. Persepsi keadilan dan transparansi berperan sebagai mediator terhadap kepatuhan dan capaian. Implementasi tetap layak pada konteks konektivitas terbatas tanpa pelanggaran privasi yang bermakna. Disimpulkan bahwa Kontrak Kelas AI efektif sebagai instrumen kebijakan mikro yang mengaitkan prinsip human-centred dengan praktik TEFL dan EAP, serta menghasilkan paket replikasi berupa kontrak, templat, dan rubrik pemeriksaan fakta untuk peningkatan mutu dan akuntabilitas pembelajaran.

Kata Kunci

Integritas Akademik; TEFL; EAP; Kontrak Kelas AI; Analitik Pembelajaran

DOI: https://doi.org/10.55480/saluscultura.v5i2.477

1. Introduction

Across English language education, the promise of a "digital nexus" for teaching and assessment now meets the pragmatics of classroom uptake, especially in TEFL and EAP courses where writing, iterative feedback, and source use sit at the core of curricular work. Sector analyses register that adoption of AI is proceeding "mainly without systematic oversight and regulation," creating a governance lag between institutional aspiration and everyday pedagogy (OECD, 2024: 6). International guidance responds by insisting that education systems "support the planning of appropriate regulations, policies and human capacity development" so that AI use "genuinely benefits and empowers teachers, learners and researchers" (UNESCO, 2023: 4). The same document centers a human-centred posture that foregrounds "human agency, inclusion, equity, gender equality, cultural and linguistic diversity" as design criteria rather than afterthoughts (UNESCO, 2023: 4). Earlier critical work set the tone with the observation that AI "is accelerating, permeating every aspect of our lives," and pressed education to ask whether diffusion would occur "without proper debate or control" (Luckin et al., 2016: 39). In TEFL/EAP classrooms, those debates become concrete decisions about what learners must disclose when AI tools assist drafting, how verification routines are enacted to prevent fabricated citations, and where boundaries are drawn in high-stakes assessment so that human judgement remains constitutive of learning. These questions sharpen in multilingual, bandwidth-constrained contexts where access is uneven, teacher workload is high, and privacy practices are variably codified, making course-level mechanisms for transparency, integrity, and data minimization a practical necessity rather than a policy luxury.

The empirical base for AI-mediated feedback in second-language writing has advanced and is no longer confined to proof-of-concept demonstrations, but the signals are conditional. A multi-level meta-analysis synthesizing prior trials reports that "overall, results... show a medium effect (g = 0.55) of automated feedback on students' writing performance," suggesting consequential gains under appropriate task conditions (Fleckenstein et al., 2023: 1). Complementing the synthesis with causal evidence, a randomized study with Chinese EFL learners concludes that "Artificial Intelligence-based instructional programs, specifically AWE, hold the potential to effectively enhance second language writing skills, especially among learners with lower proficiency levels" (Wei et al., 2023: 1). At the same time, practitioner-facing syntheses caution that guidance remains underdeveloped: "the current research and guidelines are limited and there is a pressing need for more comprehensive investigation" into AI use in ELT systems (British Council, 2024: 9). Mapping the wider higher-education terrain, a large-scale review identifies "four areas of application of AI in higher education: (1) profiling and prediction, (2) intelligent tutoring systems, (3) adaptive systems and personalization, and (4) assessment and evaluation," a distribution that helps explain both enthusiasm for timely feedback and anxiety about opacity, validity, and bias (Zawacki-Richter et al., 2019: 1). For TEFL/EAP, the same distribution translates into automated writing evaluators, conversational agents, and adaptive practice tools juxtaposed with authenticity checks and evolving assessment regimes. Without explicit classroom protocols, AI assistance risks drifting toward invisible mediation of text, weak source verification, and inconsistent expectations, while data-hungry tools raise practical questions about what information is collected, how long it is retained, and who can access it. These tensions make a case for course-embedded instruments that render AI use inspectable, align formative support with genre-specific goals, and create a record of the human work—planning, sourcing, revising—that instruction aims to cultivate.

Against this backdrop, the present study frames its problem as the translation of macro-level principles into micro-level, auditable practices that are feasible for teachers and legible to students in TEFL/EAP. The approach operationalizes human-centred guidance through a concise classroom contract that requires per-assignment disclosure of tools used, prompt rationales, and verification steps; constrains AI in high-stakes assessments; and codifies human-in-the-loop feedback so that model outputs remain subordinate to pedagogical intent. The contract is paired with lightweight artefacts—disclosure templates, fact-check rubrics, and worked examples of good and poor practice—alongside learning-management nudges that normalize transparency without inflating administrative burden. Evaluation focuses on process indicators such as completeness of transparency statements, integrity incidents involving mis-citation or fabrication, depth of revision measured through draft cycles, time on task, and punctuality, together with performance on writing and speaking rubrics and student perceptions of fairness and transparency. In bandwidth-limited contexts, provisions for offline-first or low-data capture accompany the design so that equitable participation does not depend on premium connectivity. Rather than predetermine outcomes, the study sets out to examine under what conditions AI-mediated feedback can be aligned with academic voice and source use, and how explicit transparency scaffolds can clarify authorship, strengthen verification, and reduce ambiguity about permissible assistance. By situating the inquiry within established policy anchors and emerging empirical signals, the investigation aims to articulate classroom-level mechanisms that can be scrutinized, iterated, and, where appropriate, adopted across TEFL/EAP courses in ways that remain accountable to learners, teachers, and institutional standards.

2. Methods

This study adopts a pragmatic, mixed-methods design to develop and test a human-centred "AI Transparency Protocol" for TEFL/EAP courses while aligning with international guidance on safety, inclusion, and data protection. Ethical guardrails are anchored in UNESCO's guidance that aims to "support the planning of appropriate regulations, policies and human capacity development" and to ensure a human-centred vision of GenAI in education (UNESCO, 2023: 4). The intervention is implemented at course level through a concise class contract covering permitted uses of GenAI, disclosure requirements (tool, prompt, verification), and data-minimization practices. The research spans eight months of one academic term plus analysis, with staged activities: co-design with instructors and student representatives; instructor training; controlled implementation across multiple intact EAP sections; and iterative refinement based on learning analytics and qualitative feedback. Outcomes target integrity (mis-citation/fabrication incidents), academic performance on genre-specific writing tasks, and student self-regulation and perceived fairness.

Sampling uses intact classes to preserve ecological validity; where feasible, sections are pair-matched on proficiency and instructor experience, with one adopting full protocol and another retaining business-as-usual disclosure. Baseline covariates include prior GPA, English proficiency, and connectivity constraints to address equity. Procedurally, all students receive a 45-minute micro-module on ethical AI use and contract expectations; instructors receive a facilitation guide, scenario bank, and a lightweight audit form that logs any GenAI support used by students at draft or revision stages. To reduce administrative burden, the audit embeds into normal submission workflows in the LMS and offers offline-first capture for low-bandwidth contexts. Integrity incidents are operationalized with reference to widely adopted definitions of third-party outsourcing—"Contract cheating happens when a third party completes work for a student who then submits it… as their own" (QAA, 2022: 3)—and detected via rubric-based review, trace audits, and targeted viva checks on a random subset.

Quantitative analysis estimates intention-to-treat and per-protocol effects on writing quality, integrity incidents, and self-regulation. Given classroom nesting, estimates include class fixed effects and robust clustering; when section counts permit, multilevel models will be reported as sensitivity. Mechanism testing follows contemporary mediation logic: mediation analysis is used "to understand, explain, or test a hypothesis about how or by what process or mechanism a variable X transmits its effect on Y," with a mediator "causally located between X and Y" (Igartua & Hayes, 2021: 1). The prespecified model evaluates whether protocol adoption (X) improves transparent process reporting and self-regulation (M), which in turn reduces integrity incidents and elevates performance (Y). Bootstrap confidence intervals quantify indirect effects; heterogeneity by bandwidth access and prior AI familiarity is examined through moderators.

Qualitative evidence triangulates mechanisms and acceptability through focus groups (students, instructors) and prompted reflections attached to submissions. Analysis employs thematic analysis, treating it as "a research method used to identify and interpret patterns or themes in a data set" with an explicit, six-step, auditable pathway from familiarization to conceptual modeling. Coding is reflexive and memo-driven, with discrepant cases sought to probe policy fatigue and unintended consequences. Throughout, governance mirrors UNESCO's emphasis on privacy and risk management, including minimized personal data, consented usage, and transparent communication of purposes to participants (UNESCO, 2023: 23). Collectively, the design balances rigor and practicality, foregrounding explainable procedures that programs can adopt without prohibitive overhead.

3. Results and Discussion

The presentation of findings is organized in two interlocking parts to preserve continuity between measurement and interpretation. The first part, Process and Performance Effects (T1–T3 Trajectories), traces changes across the three collection points for core indicators: completeness of AI-use disclosures, integrity incidents, revision intensity, time-on-task, punctuality, and rubric-based scores for genre-specific writing and speaking. Descriptive patterns are established and then examined with class-aware models that report effect sizes, confidence intervals, and robustness checks, allowing time trends to be read alongside contrasts between sections. The second part, Mechanisms, Equity, and Micro-governance Implications, interrogates how perceived fairness and transparency may transmit effects to compliance and outcomes through mediation analyses, while heterogeneity by bandwidth constraints and prior AI familiarity is assessed to surface equity-relevant differences. Qualitative materials from focus groups and prompted reflections are integrated to clarify mechanisms and contextual contingencies. Together, these parts move from trajectories to explanations without presuming conclusions in advance.

3.1. Process and performance effects (T1-T3 trajectories)

The analysis of process and performance unfolds as a continuous trajectory across three scheduled moments of measurement: T1 as baseline, T2 as mid-course monitoring, and T3 as endline. The class-embedded protocol asked students to disclose any AI support used during drafting or revision, to record brief verification steps for sources and claims, and to observe bounded-use parameters in high-stakes assessments. Framed by international guidance that urges education systems to "support the planning of appropriate regulations, policies and human capacity development" so that AI use "genuinely benefits and empowers teachers, learners and researchers" (UNESCO, 2023: 4), the narrative that follows examines whether such transparency scaffolds moved from policy rhetoric into the mundane routines of classroom work. The aim is to read descriptive movement in indicators alongside class-aware estimation while drawing on the literature to interpret mechanisms and to avoid over-generalization.

From T1 to T3, completeness of AI-use disclosures rose in a pattern that was steepest between the first two measurements. Early reflections suggested that students initially hesitated about what counted as assistive versus substitutive use, but that hesitation waned as disclosure templates and worked exemplars clarified expectations. This development is coherent with the policy turn toward concrete, auditable classroom artifacts; UNESCO's human-centred vision explicitly foregrounds "human agency, inclusion, equity, gender equality, cultural and linguistic diversity" as operational criteria rather than afterthoughts (UNESCO, 2023: 4). The British Council's observation that "the current research and guidelines are limited and there is a pressing need for more comprehensive investigation" (British Council, 2024: 9) also helps explain the unevenness at baseline and the subsequent convergence once routine fields for tool names, prompt rationales, and verification steps became part of the submission grammar.

Incidents that threatened integrity—particularly mis-citation and fabrication detected through rubric-based audits and selective vivas—declined across the term, with the sharpest contraction again visible by T2 and consolidation by T3. The decline did not appear to stem from punitive surveillance; rather, it reflected greater clarity about permissible assistance and the normalization of rapid source checks in AI-supported drafts. A definitional touchstone helped stabilize that clarity: "Contract cheating happens when a third party completes work for a student who then submits it ... as their own, where such input is not permitted" (QAA, 2022: 3). Although the course focus was disclosure rather than outsourcing, the cited boundary gave instructors a shared language for explaining why transparency matters for assessment validity and gave students a reasoned basis for distinguishing formative assistance from prohibited substitution.

Revision intensity increased steadily from T1 to T3 when counted as meaningful iterations that altered structure, stance, or evidence selection. This shift aligned with feedback designs that required students to log prompts, justify accept/reject decisions for model suggestions, and show how revisions advanced disciplinary voice. The pattern accords with contemporary feedback scholarship: "feedback practice should place less emphasis on what teachers do ... and more emphasis on how students generate, make sense of, and use feedback" (Winstone & Carless, 2020: 17). As the locus of effort moved from receipt to uptake, revisions clustered around organization, claim support, and audience alignment rather than only surface correctness, indicating that transparency scaffolds were not merely procedural but also epistemic, encouraging learners to articulate the relationship between suggestions and genre expectations.

Performance on genre-specific writing rubrics also rose over time, with particularly visible gains for lower-proficiency cohorts and for tasks that required argument structure and source integration. Two strands of evidence temper and contextualize these trajectories. A multi-level meta-analysis reported that "overall, results ... show a medium effect (g = 0.55) of automated feedback on students' writing performance," while cautioning that automated feedback is not a single, uniform intervention and must be conditioned by task design (Fleckenstein et al., 2023: 1). A randomized trial similarly concluded that AI-based writing programs "hold the potential to effectively enhance second language writing skills, especially among learners with lower proficiency levels" (Wei et al., 2023: 1). The observed movement from T1 to T3 is compatible with these signals insofar as the protocol channeled AI support through verification and authorship boundaries, steering improvement toward organization and evidence rather than unreflective text substitution.

Speaking performance—short academic talks with brief Q&A—showed more modest but consistent gains, particularly in coherence and discourse marking. Although AI support was less direct for oral work, the writing transparency routine migrated to speaking preparation: students commonly used AI to outline and to anticipate questions, then annotated how claims had been validated with sources. A broader cartography of AI in higher education helps situate this behavior; a large-scale review synthesized "four areas of application of AI in higher education: (1) profiling and prediction, (2) intelligent tutoring systems, (3) adaptive systems and personalisation, and (4) assessment and evaluation" (Zawacki-Richter et al., 2019: 1). The observed improvements appear to align with adaptive drafting support and tutor-like rehearsal prompts, filtered through rules that prioritize authorship and verification rather than performance mimicry.

Time-on-task increased from T1 to T2 for most learners and then plateaued or slightly declined by T3 as drafting and verification routines became more efficient. Such temporal metrics warrant caution. As a methodological reminder, "time-on-task measures are used to provide a more 'accurate' estimate of student learning," yet the "adoption of particular time-on-task estimation strategy can have a significant effect on the overall fit of the model" (Kovanović et al., 2015: 1; 9). To avoid over-weighting noisy estimates, temporal logs were read in tandem with product-oriented indicators such as revision depth and rubric sub-scores. The combined picture suggests that transparency initially extends engagement as learners practice verification, then compresses once strategies are internalized and more quickly deployed.

Punctuality improved notably between T1 and T2 and held through T3. Across assignments, timelier submission clustered with two process markers: early completeness in disclosure and explicit plan-of-work notes logged at the outset of drafting. The co-movement is consistent with self-regulation accounts in which planning and monitoring underpin task management; "self-regulated learning ... includes the cognitive, metacognitive, behavioral, motivational, and emotional/affective aspects of learning" (Panadero, 2017: 1). The protocol's short planning prompts appear to have externalized part of this regulation, distributing effort more evenly across drafting cycles and reducing last-minute surges that correlate with integrity lapses.

Contrasts between intact sections illuminated dose and practice effects. Classes that adopted the full protocol earlier showed higher T2 disclosure completeness and fewer integrity incidents than classes with delayed adoption; by T3, gaps narrowed but did not disappear. Facilitation logs indicated more frequent reference to exemplars and quicker feedback on incomplete disclosures in early-adopter sections. The British Council's remark that practitioner guidance is thin—"current research and guidelines are limited ..." (British Council, 2024: 9)—helps explain why effects clustered around sections with abundant practical examples and repeated modeling. Where contract details arrived late, students took longer to form stable habits, and instructors flagged more borderline cases requiring clarification.

Equity-relevant heterogeneity was visible but tractable. Learners facing bandwidth constraints registered slower growth in revision intensity at T2 but converged by T3 once offline-first templates and printable checklists were emphasized. Prior AI familiarity predicted early completeness in disclosures but had little bearing on endline performance once verification routines were mastered. Such patterns echo the human-centred emphasis on inclusion and linguistic diversity, which calls for designs robust to uneven access; the guidance's insistence on a "human-centred approach" is not only a slogan but a design requirement (UNESCO, 2023: 4). The convergence by term's end suggests that small, connectivity-aware choices—prompt banks that do not require persistent access and minimized data capture—can mitigate initial disparities without diluting academic expectations.

Genre-specific analysis offers a clearer view of where gains concentrate. In research-style reports, organization and evidential support improved more than style metrics, plausibly because verification routines targeted citation chains and claim-evidence alignment. In position papers, stance and hedging advanced as students practiced translating AI-suggested phrasing into discipline-appropriate voice. These shifts are consonant with the meta-analytic caution that "the use of automated feedback tools cannot be understood as a single consistent form of intervention" (Fleckenstein et al., 2023: 1); rather, effects depend on how feedback is framed and acted upon. The class contract and its associated artifacts appear to have supplied the mediating structure necessary for improvement where pedagogy needed it most.

Coupling between process and product tightened across the term. Assignments with higher disclosure completeness and deeper revision cycles tended to earn stronger rubric outcomes, suggesting that transparency scaffolds were not only compliance devices but also catalysts for more deliberate drafting. While causal claims must remain modest outside the planned comparisons, the association complements controlled evidence that AWE "hold[s] the potential to effectively enhance second language writing skills" under structured conditions (Wei et al., 2023: 1) and coheres with feedback-literacy perspectives that position learners to generate, interpret, and use feedback rather than receive it passively (Winstone & Carless, 2020: 17). In practice, transparency made thinking visible, and visible thinking supported more targeted instruction and self-correction.

Concerns about assessment validity diminished as disclosure quality improved. Instructors reported that incomplete disclosures at T1 complicated judgments about sudden fluency spikes or atypical phrasing. By T3, with fuller process records, viva checks shifted from suspicion toward scholarly conversation about source trails and reasoning steps, and grading decisions reflected greater confidence. The definitional clarity provided by integrity frameworks—such as the QAA formulation that "contract cheating happens when a third party completes work for a student ... where such input is not permitted" (QAA, 2022: 3)—served as a backdrop rather than a cudgel, enabling the maintenance of firm boundaries without defaulting to blanket bans that might suppress legitimate, scaffolded tool use in a second-language setting.

Taken together as a narrative from T1 through T3, the trajectories depict modest but durable reconfiguration of process indicators under a transparency-first regimen: disclosures became more complete, integrity incidents declined, revisions deepened, time on task stabilized at a more efficient level, and punctuality improved, with performance gains concentrated in organization and evidence. These movements sit within a wider map of higher-education applications that encompasses "assessment and evaluation," "adaptive systems and personalisation," and "intelligent tutoring systems" (Zawacki-Richter et al., 2019: 1), but they do so in a form that makes AI mediation inspectable and accountable to classroom aims. They also remain bounded by methodological cautions about temporal analytics—"adoption of particular time-on-task estimation strategy can have a significant effect on the overall fit of the model" (Kovanović et al., 2015: 9)—and by the variability of tools and tasks across contexts. The subsequent analysis attends to mechanisms, asking whether perceived fairness and transparency transmit these observed changes to compliance and outcomes, while remaining attentive to equity-relevant differences that accompany bandwidth, linguistic repertoire, and prior familiarity with AI.

3.2. Mechanism, equity, and micro-governance implication.

The operative mechanisms linking a transparency-first protocol to changes in process and performance can be framed as the combined action of disclosure, verification, and bounded use on learners' feedback literacy and self-regulation, under conditions that respect human agency and inclusion. A human-centred anchor is essential, not incidental. International guidance emphasizes that educational systems should "support the planning of appropriate regulations, policies and human capacity development" so that AI adoption "genuinely benefits and empowers teachers, learners and researchers" (UNESCO, 2023: 4). In practical terms, a class-level contract translates that injunction into routines that make AI mediation visible: naming the tool, recording the prompt rationale, and documenting the verification pathway for cited claims. Visibility invites evaluability. Once assistance is inspectable, teachers can calibrate how much support is educationally appropriate for the genre at hand, and students can rehearse the metacognitive moves that convert suggestions into disciplinary argument. The shift is consonant with contemporary feedback theory: "feedback practice should place less emphasis on what teachers do ... and more emphasis on how students generate, make sense of, and use feedback" (Winstone & Carless, 2020: 17). By requiring brief, auditable explanations of what was accepted, modified, or rejected, the protocol repositions feedback as learner action rather than one-way transmission.

A second mechanism concerns clarity of boundaries and its downstream effect on integrity. In the absence of explicit rules, students are left to infer where help ends and substitution begins, especially when large language models produce fluent text. Definitional clarity helps, not as punishment but as guidance. The sector's integrity benchmark states that "Contract cheating happens when a third party completes work for a student who then submits it ... as their own, where such input is not permitted" (QAA, 2022: 3). A classroom contract operationalizes this line by distinguishing formative drafting support (permitted with disclosure) from prohibited replacement of authorship (not permitted), and by pairing the distinction with brief viva prompts or fact-check rubrics that normalize verification. The result is a governance mechanism that reduces ambiguity at the point of assessment, thereby supporting validity without outlawing legitimate, scaffolded AI use in second-language contexts. This orientation aligns with the broader caution that institutional adoption is proceeding "mainly without systematic oversight and regulation" (OECD, 2024: 6), a situation that heightens the value of micro-governance capable of operating inside courses rather than waiting for system-wide regulation.

A third mechanism is the coupling of transparency with self-regulated learning. Planning the drafting path, monitoring sources and time-on-task, and evaluating revisions are not side activities; they are the work of academic literacy. Self-regulation has been defined to encompass "the cognitive, metacognitive, behavioral, motivational, and emotional/affective aspects of learning" (Panadero, 2017: 1). Short planning prompts and disclosure fields externalize parts of this regulation, turning private decisions into small public commitments that can be coached. When students articulate why a model's suggestion serves the argument, how evidence has been checked, and what remains to be revised, the proximate product is a cleaner audit trail; the deeper product is a practiced sequence of strategic behavior. That sequence also clarifies responsibility for textual choices, strengthening authorship and reducing the risk that fluent but unexamined output stands in place of disciplinary reasoning.

These mechanisms interact with measurement in non-trivial ways. Time-on-task, a staple of learning analytics, is notoriously sensitive to estimation choices; "time-on-task measures are used to provide a more 'accurate' estimate of student learning," yet the "adoption of particular time-on-task estimation strategy can have a significant effect on the overall fit of the model" (Kovanović et al., 2015: 1; 9). Micro-governance that foregrounds verification and revision depth provides alternate anchors—product indicators and process justifications—so that temporal logs can be interpreted against substantive evidence of learning activity rather than treated as proxies. In the same spirit, writing quality gains under automated support require contextualization. A multi-level meta-analysis found that "overall, results ... show a medium effect (g = 0.55) of automated feedback on students' writing performance," while cautioning that tools do not constitute "a single consistent form of intervention" (Fleckenstein et al., 2023: 1). A randomized study likewise reported that AI-based programs "hold the potential to effectively enhance second language writing skills, especially among learners with lower proficiency levels" (Wei et al., 2023: 1). Micro-governance matters here because it steers the locus of effect: the same tool can nudge toward surface optimization or toward structure and evidence, depending on whether disclosure and verification are codified as part of the task.

Equity considerations shape how these mechanisms travel across learners and settings. Policy documents insist on a human-centred posture that foregrounds "human agency, inclusion, equity, gender equality, cultural and linguistic diversity" (UNESCO, 2023: 4), and system analyses warn that uncoordinated diffusion risks amplifying disparities unless adoption is paired with capacity-building and accountability (OECD, 2024: 6–7). In bandwidth-constrained environments, transparency routines cannot presume persistent connectivity or device homogeneity. Offline-first templates, printable checklists, and minimized data capture become more than conveniences; they are equity levers. These small design choices support parity of participation by allowing students to prepare disclosures and verification notes without streaming access and by ensuring that evidence of process does not require data-intensive tools. Equity also touches language repertoires. Where students write in a second or third language, the explicit separation of idea formation, language support, and source verification helps prevent the conflation of linguistic scaffolding with intellectual substitution. In multilingual classrooms, such separation protects the learner's epistemic agency—what counts as their claim, their warrant, their evidence—while allowing modest, disclosed assistance with phrasing when appropriate.

Teacher capacity is a companion axis of equity. Without concrete artifacts and exemplars, instructors face the double burden of policing misuse and inventing new assessment routines on the fly. Practitioner syntheses concede that "the current research and guidelines are limited and there is a pressing need for more comprehensive investigation" (British Council, 2024: 9). Micro-governance reduces this burden by providing ready-to-adapt templates, short scripts for classroom discussion of permissible assistance, and audit checklists that fit within existing workflows. Importantly, the artifacts must be legible and light. If compliance imposes heavy administrative load, policy fatigue grows and students learn the wrong lesson—that transparency is a hoop rather than a habit. Guidance from assessment communities gestures toward this balance: authentic assessments can "allow use with acknowledgement" and can include "assessment of the process/use of GenAI," thereby realigning incentives toward openness rather than concealment (Jisc, 2024: 4). Instructors, in turn, gain a clearer basis for judgment because process records make authorship and source claims discussable at the level of evidence, not suspicion.

From a governance perspective, the micro layer does not replace institutional policy; it informs it. AIEd remains a moving target, with a widely cited review noting "four areas of application of AI in higher education: (1) profiling and prediction, (2) intelligent tutoring systems, (3) adaptive systems and personalisation, and (4) assessment and evaluation" (Zawacki-Richter et al., 2019: 1). Course-embedded protocols supply empirically tractable practices within the latter three areas, allowing departments to observe what works in context before codifying broader rules.

This sequencing reflects a precautionary pragmatism: rather than prohibit or mandate wholesale, institutions can authorize locally auditable trials that report on transparency completeness, integrity incidents, and genre-specific outcomes. Governance thus proceeds by disciplined iteration, not by technology-first decree. The ethical rationale echoes an early provocation: AI "is accelerating, permeating every aspect of our lives," and the key question is whether diffusion occurs "without proper debate or control" (Luckin et al., 2016: 39). Micro-governance is a site for that debate, precisely because it couples concrete classroom routines with data that can travel upward to inform policy.

Mechanism testing through mediation analysis clarifies how transparency might transmit benefits to outcomes. Mediation has been described as a way "to understand, explain, or test a hypothesis about how or by what process or mechanism a variable X transmits its effect on Y," with the mediator "causally located between X and Y" (Igartua & Hayes, 2021: 1). In this logic, protocol adoption (X) is expected to increase perceived fairness and transparency (M) by making assistance and verification explicit, which in turn aligns expectations and reduces the likelihood of integrity breaches that depress performance (Y). Fairness perceptions matter because students judge the legitimacy of rules not only by outcomes but by procedures—who must disclose what, how consistently policies are applied, and whether sanctions, if any, are proportionate. A classroom contract that concentrates on process reporting rather than tool bans communicates a fairness narrative: openness is rewarded, concealment is discouraged, and responsibility is shared through staged verification.

Micro-governance also addresses assessment validity, which is central to EAP/TEFL. Validity concerns arise when sudden changes in textual fluency or content density cannot be explained by observable learning processes. Process records, including disclosure completeness and revision notes, restore interpretability. They allow instructors to ask productive questions about source trails and reasoning steps rather than default to suspicion. In turn, students encounter assessment as a dialogic practice about evidence quality and argumentative coherence, not only a check for misconduct. This reorientation is consistent with the broader shift from surveillance to pedagogy in academic integrity work, a shift that gains traction when expectations are articulate and when students are equipped to meet them.

The equity implications of data handling merit explicit attention. Transparency requires some record of process, but data minimization and role-based access remain non-negotiable. A human-centred posture has to balance documentation with privacy by limiting personally identifiable information, separating identity from analysis files, and communicating the purpose and duration of data retention in plain language. Those moves are not merely legal shields; they are trust-building devices. When learners see that governance honors not just content standards but also informational rights, compliance becomes cooperation, and transparency becomes a norm rather than a fear response.

At institutional scale, the distributable artifact set—contract template, AI-use declaration form, verification checklist, example-driven briefing slides—functions as a policy starter kit. Programs can adopt, adapt, and evaluate these materials across courses, sharing descriptive indicators and qualitative themes to refine practice. Over time, these micro-level signals can inform programme-wide assessment redesigns, for example by shifting weight toward process portfolios, viva-supported submissions, or authentic tasks that are less susceptible to substitution. Guidance from assessment bodies underscores the feasibility of such reweighting; allowing use "with acknowledgement" and assessing "the process/use of GenAI" places integrity within the design of tasks rather than solely within policing after the fact (Jisc, 2024: 4). In EAP/TEFL, genre-based rubrics can integrate process criteria without eclipsing product quality, maintaining attention to organization, evidence, and voice while valuing transparent method.

Finally, micro-governance helps to reconcile innovation with caution. AIEd scholarship has catalogued wide variation in definitions and applications, observing that many educators "are unaware of its scope and, above all, of what it consists of" (Zawacki-Richter et al., 2019: 2). That uncertainty is a call for designs that teach as they govern: routines that not only demand disclosure but also model how to disclose; prompts that not only ask for verification but also exemplify it; rules that not only draw boundaries but also explain why boundaries exist. In such arrangements, mechanisms of improvement—feedback literacy, self-regulation, and perceived fairness—become integral to equity and to the legitimacy of classroom governance. Institutional policy then has a living substrate: not abstractions, but practices that have been shown to be intelligible, auditable, and adaptable in the contingent realities of TEFL/EAP teaching.

4. Conclusion

The study demonstrates that a transparency-first, human-centred classroom protocol—anchored in a concise AI use contract, brief disclosure and verification steps, and bounded use in high-stakes assessment—can realign core learning processes in TEFL/EAP without displacing pedagogic intent. Across the three measurement points, completeness of AI-use disclosures increased, integrity incidents related to mis-citation and fabrication declined, revision shifted from surface correction toward organization and evidential support, time on task stabilized after an initial learning phase, punctuality improved, and performance gains were most visible in genre-specific writing criteria, with modest improvements in speaking coherence. These patterns address the research aims by translating macro-level guidance into auditable micro-practices, by linking those practices to measurable changes in processes and outcomes, and by indicating that perceived fairness and transparency operate as plausible pathways connecting protocol adoption to compliance and academic performance. Feasibility in bandwidth-limited settings was supported through offline-first templates and lightweight artifacts, suggesting that equity can be advanced through design rather than exception. At the same time, conclusions remain bounded by intact-class sampling, sensitivity of temporal analytics, tool heterogeneity, and the self-reported elements of disclosure records. A realistic next step is multi-site replication with stronger counterfactuals, refinement of instruments to capture revision depth more reliably, and longitudinal follow-up to test persistence and transfer to additional genres. Taken together, the results recommend a governance posture that rewards transparent method and verifiable sourcing, enabling programs to harness AI's affordances while safeguarding authorship, integrity, and the long-term development of academic voice.

Acknowledgment

I would like to sincerely thank everyone who contributed to this research. My deepest gratitude is extended to the RisetMU for the financial assistance that was crucial for the completion of this project. Their support is greatly appreciated.

References

British Council. (2024). *Artificial intelligence and English language teaching: Preparing for the future*. British Council. Google Scholar

Bretag, T. (2019). Contract cheating in higher education: A comprehensive review. *Higher Education, 77*, 593–610.

Brookhart, S. M. (2017). How to give effective feedback to your students (2nd ed.). *ASCD*. Google Scholar

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. Crossref | Google Scholar

Eaton, S. E. (2021). *Plagiarism in higher education: Tackling tough topics in academic integrity*. Libraries Unlimited. Google Scholar

Fitriati, S. W. (2025). *AI-enhanced self-regulated learning among Indonesian EFL learners*. ERIC.

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence, 6*, 1162454. Crossref | Google Scholar

Hyland, K., & Hyland, F. (2019). Feedback in second language writing (2nd ed.). Routledge. Google Scholar

Igartua, J. J., & Hayes, A. F. (2021). Mediation, moderation, and conditional process analysis in communication research. *Communication Methods and Measures, 15*(3), 1–22. Google Scholar

Jisc. (2024). Embracing generative AI in assessments: A guided approach. Jisc. Google Scholar

Kasneci, S., Sessler, K., Straubinger, S., & Kasneci, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. Crossref | Google Scholar

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015, March). Penetrating the black box of time-on-task estimation. In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 184-193). Crossref | Google Scholar

Liu, S., Zhang, S., & Dai, Y. (2025). Do mobile games improve language learning? A meta-analysis. *Computer Assisted Language Learning*, 1-29. Crossref | Google Scholar

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. *Pearson*. Google Scholar

Mah, C. (2024). Teacher versus LLM feedback: Qualitative insights with randomized evidence context. Annenberg Institute EdWorkingPapers.

Meyer, J., Nguyen, T., Wang, L., & Smith, A. (2024). Using large language models to bring evidence-based feedback into the classroom. *Computers and Education: Artificial Intelligence, 5*, 100187. Google Scholar

Nicol, D. (2020). The power of internal feedback: Exploiting the social, cognitive and evaluative processes of self-assessment. *Assessment & Evaluation in Higher Education, 45*(5), 1–12. Crossref | Google Scholar

OECD. (2024). The potential impact of AI on equity and inclusion in education. OECD Publishing.

Pack, A. (2024). Using AI in TESOL: Ethical and pedagogical considerations. TESOL Quarterly.

Pan, Z., Xing, W., & Chen, G. (2024). Learning analytics interventions in higher education: A systematic review. *Journal of Learning Analytics, 11*(2), 1–28.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422. Crossref | Google Scholar

Peña-Acuña, B., López-Meneses, E., & Vázquez-Cano, E. (2024). *Learning English with AI: Opportunities and challenges*. Frontiers in Education.

QAA. (2022). *Contracting to cheat in higher education: How to address contract cheating*. The Quality Assurance Agency for Higher Education (UK). Google Scholar

Qiu, X., Ma, Y., & Chen, H. (2024). The effects of virtual reality on EFL learning: A meta-analysis. *Education and Information Technologies*. Crossref | Google Scholar

TESOL International Association. (2024). *Position statement on artificial intelligence in TESOL*. TESOL.

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.

Werdiningsih, I. (2024). *Strategic use of ChatGPT by postgraduate EFL students: Perceptions and practices*. Cogent Education.

Wei, P., Wang, X., & Dong, H. (2023). *Artificial intelligence-based writing instruction for EFL students: A randomized controlled trial*. Frontiers in Psychology.

Winstone, N. E., & Carless, D. (2020). *Designing effective feedback processes in higher education: A learning-focused approach*. Routledge. Google Scholar

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education, 16*(39), 1–27. Crossref | Google Scholar