

Title: "I Don't Know Where My Ability Stands": When AI-Disclosure Mechanisms Backfire in Exam-Centered EFL Reading

Research Overview: This qualitative case study examines how five Grade 12 Vietnamese EFL learners, whose reading training transitioned from PDF-based tasks to AI-assisted reading from November 2025 to April 2026, perceive an AI-integrated Gemini reading platform and navigate voluntary external AI consultation during high-stakes IELTS preparation. Notably, IELTS scores are increasingly used by Vietnamese universities for direct admission and admission bonuses. A dual-layer TAM framework yields a three-way integrity typology and two emergent insights challenging prevailing assumptions about AI-disclosure mechanisms in language learning.

Purpose of Study: The diffusion of generative AI into language classrooms has outpaced theoretical understanding of how learners simultaneously perceive a teacher-mandated AI platform and decide, autonomously, whether to consult external AI tools (e.g., ChatGPT, Gemini) within the same task cycle (Landa-Blanco, 2025; Nguyen, 2025; Nguyen Thi et al., 2025; Vo & Nguyen, 2024). Existing Technology Acceptance Model (TAM) research in computer-assisted language learning (CALL) treats the platform as the unit of analysis and does not distinguish mandated platform use from learner-initiated AI consultation, nor does it engage substantively with academic integrity reasoning. This study addresses both gaps through a dual-layer analytical framework separating (1) learner perception of a teacher-assigned AI-integrated reading platform (Layer 1) from (2) learner-initiated external AI consultation behaviour during and after the task (Layer 2). Anchored in Cognitive Fit Theory (Vessey, 1991), TAM (Davis, 1989), UTAUT2 (Venkatesh et al., 2012), and Self-Determination Theory (Ryan & Deci, 2000) as a background motivational lens, alongside AI-acceptance constructs from Huang et al. (2025), the study asks: how do Vietnamese EFL learners (IELTS band 5.5-7.0) perceive the design features of an AI-integrated reading platform, and how do their motivational orientations and integrity positions shape their external AI consultation behaviour? Vietnam—where high-stakes IELTS preparation for tertiary admission, rapid generative-AI uptake, and exam-centered academic norms intersect—is treated not as a sample but as analytical leverage on AI-policy questions facing global education.

Research Methods: Five Grade 12 students enrolled in the same IELTS preparation class completed identical mandatory reading tasks on the same teacher-designed Gemini platform under one teacher's instruction. Holding teachers, platform, and task constant produces a bounded case in which motivational orientation and integrity position can be compared without major instructional confounds. All five participated in semi-structured interviews (31–63 minutes/person) conducted in Vietnamese, structured around the framework's eight construct areas. Data were analysed through hybrid deductive-inductive coding combining directed content analysis (Hsieh & Shannon, 2005) with the hybrid procedure of Fereday and Muir-Cochrane (2006), following Saldaña (2021). Deductive codes were organised under nine framework-derived construct groups; 12 additional inductive codes emerged through iterative close reading. Three researchers coded via manual comment-bubble annotation, with regular consensus meetings resolving discrepancies. Intercoder reliability procedures are in progress on one full transcript; Cohen's κ will be reported in the full paper. Analytical claims are advanced as exploratory hypotheses grounded in cross-case analysis. Layer 2 (AI as comprehension partner) was prominently evident in only one participant; the remaining four used external tools more narrowly—primarily for vocabulary lookup or dictionary-style checks—indicating that comprehension-level engagement remains relatively rare even when external AI tools are available within this exam-driven cohort.

Results and Discussion: Three findings are advanced. **First**, all five participants preferred the platform's vocabulary list to appear after rather than before reading, but offered five theoretically distinct rationales—spoiler prevention, self-assessment preservation, frequency-based acquisition, comprehension calibration, and natural retention. A single design preference can therefore arise from incompatible learning logics, complicating cognitive-fit assumptions of a single optimal user model. **Second**, cross-participant analysis generated a three-way academic integrity typology not anticipated by the dual-layer framework: *rule-based*, framing integrity as a temporal boundary between during-task and post-task AI use; *epistemic*, framing integrity as preserving accurate self-knowledge of one's reading ability and authentic training trajectory; and *consequentialist*, framing integrity in terms of long-term skill consequences. The three orientations are advanced as analytical categories for hypothesis generation and future instrument development. **Third**, two inductive findings invert standard AI-transparency assumptions. (a) The *Shame-Demotivation*

Hypothesis: mandatory AI-disclosure mechanisms could induce embarrassment and erode learning motivation under examination pressure, reversing the standard claim that disclosure deters misuse.

(b) The *Báo Động (Alarm Signal) Argument*: habitual disclosure normalises AI consumption, removing guilt without resolving the underlying skill gap; score-reporting pressure was further identified as a structural driver of misuse that operates independently of (or in addition to) individual integrity values. Together, these findings indicate that platform-level transparency policies may produce unintended psychological and systemic effects that current TAM instruments are not designed to capture. Whether such effects are more readily observable in exam-centred settings than in less test-driven contexts remains a hypothesis for future comparative research.

Keywords: AI-assisted EFL reading; academic integrity typology; AI-disclosure mechanisms; dual-layer TAM; exam-centered EFL settings