

**Humanities and Educational
Sciences Journal**

ISSN: 2617-5908 (print)



**مجلة العلوم التربوية
والدراسات الإنسانية**

ISSN: 2709-0302 (online)

**Human-AI Collaboration in Language Assessment:
EFL Learners' Trust, Anxiety, and AI Literacy in
AI-Mediated Writing Feedback(*)**

<https://doi.org/10.55074/hesj.vi57.1998>

Nadia Abdullah Alshahrani

Assistant Professor, English Language Institute
Princess Nourah bint Abdulrahman
University, Riyadh, KSA

*) Received: 8/5/2026

Accepted: 9/7/2026

) Journal Website:

<http://hesj.org/ojs/index.php/hesj/index>

العدد (57)، شهر نوفمبر 2026 م
Issue (57) November 2026

897

مجلة العلوم التربوية والدراسات الإنسانية
Humanities and Educational Sciences Journal

Human-AI Collaboration in Language Assessment: EFL Learners' Trust, Anxiety, and AI Literacy in AI-Mediated Writing Feedback

Dr. Nadia Abdullah Alshahrani

Assistant Professor, English Language Institute

Princess Nourah bint Abdulrahman

University, Riyadh, KSA

Abstract

This study aimed to examine how EFL learners' trust in AI feedback, AI-related anxiety, and AI literacy relate to their engagement with, and uptake of, AI-mediated written corrective feedback. An explanatory sequential mixed-methods design combined survey data from 300 university EFL learners, analysed with partial least squares structural equation modelling (PLS-SEM), with follow-up semi-structured interviews ($n = 16$); the structural model explained 48% of the variance in feedback uptake. AI literacy positively predicted trust ($\beta = .46$) and negatively predicted anxiety ($\beta = -.39$); trust was the strongest positive predictor of both engagement ($\beta = .41$) and uptake ($\beta = .44$), while anxiety exerted smaller negative effects. A cluster analysis identified three learner profiles - Confident Collaborators, Anxious Verifiers, and Detached Low-Literacy Users. Because anxiety was measured at three points across the period of AI-mediated collaboration, change over time was also examined: a repeated-measures ANOVA showed AI anxiety declining significantly across the three waves (partial $\eta^2 = .18$), and a latent growth curve model showed that higher AI literacy and higher trust each predicted a steeper decline. The qualitative findings showed that learners did not accept AI feedback uncritically but practised calibrated trust, verifying machine feedback against teachers and peers, and that AI literacy functioned as a regulating resource that tempered both over-reliance and anxiety. On this basis the study proposes a Calibrated Trust Framework for AI-mediated assessment, in which AI literacy supports calibrated trust, calibrated trust drives engagement, and engagement converts into uptake. The study advances a learner-centred account of human-AI collaboration in language assessment and argues that AI literacy, rather than tool sophistication alone, is central to responsible, pedagogically meaningful, AI-mediated assessment. Implications for assessment design, teacher education, and ethical practice are discussed.

Keywords: AI-mediated feedback; trust; AI anxiety; AI literacy; language assessment; EFL writing.

نادية بنت عبد الله الشهراني
أستاذ تدريس اللغة الإنجليزية للناطقين بغيرها المساعد
معهد اللغة الإنجليزية، جامعة الاميرة نورة بنت عبد الرحمن
الرياض - السعودية

المستخلص:

هدفت هذه الدراسة إلى فحص العلاقة بين ثقة متعلمي اللغة الإنجليزية بوصفها لغةً أجنبية بالتغذية الراجعة المقدمة بالذكاء الاصطناعي، وفلقتهم المرتبط به، ومستوى إلمامهم به، وبين انخراطهم في التغذية الراجعة التصحيحية الكتابية المدعومة بالذكاء الاصطناعي واستفادتهم منها. واعتمدت الدراسة تصميمًا تنابعيًا تفسيريًا للطرائق المختلطة، جمع بين بيانات استبانة من 300 متعلم جامعي للغة الإنجليزية لغةً أجنبية، خللت باستخدام نموذج المعادلات الهيكلية بالمربعات الصغرى الجزئية (PLS-SEM)، ومقاييل شبة مقننة لاحقة مع 16 مشاركا. وفُسر النموذج الهيكلي 48% من التباين في الاستفادة من التغذية الراجعة.

تنبأ الإلمام بالذكاء الاصطناعي إيجابيًا بالثقة ($\beta = .46$)، وسلبًا بالقلق ($\beta = -.39$) وكانت الثقة أقوى متنبئ إيجابي بكل من الانخراط ($\beta = .41$) والاستفادة ($\beta = .44$)، في حين كان للقلق أثر سلبي أقل قوة. وكشف التحليل العنقودي عن ثلاثة أنماط من المتعلمين: المتعاونون الواثقون، والمتحققون القلقون، والمستخدمون المنفصلون ذوو الإلمام المحدود.

ولأن القلق من الذكاء الاصطناعي قيس في ثلاث نقاط زمنية خلال فترة التعاون المدعوم بالذكاء الاصطناعي، فُحص التغير فيه بمرور الوقت أيضًا. وأظهر تحليل التباين للقياسات المتكررة انخفاضًا دالًا إحصائيًا في القلق عبر الموجات الثلاث (η^2 الجزئي = .18). (كما يتنبأ نموذج منحني النمو الكامن أن ارتفاع مستوى الإلمام بالذكاء الاصطناعي والثقة به، كلٌّ على حدة، تنبأ بانخفاض أشد في القلق بمرور الوقت.

وأظهرت النتائج النوعية أن المتعلمين لم يقبلوا ملاحظات الذكاء الاصطناعي دون تمحيص، بل مارسوا ثقةً متوازنة، فتحققوا من ملاحظات الآلة بمقارنتها بملاحظات المعلمين والأقران. كما أدى الإلمام بالذكاء الاصطناعي دور مورد تنظيمي يحد من الإفراط في الاعتماد عليه ومن القلق تجاهه. وبناءً على ذلك، تقترح الدراسة «إطار الثقة المتوازنة في التقييم المدعوم بالذكاء الاصطناعي»، الذي يدعم فيه الإلمام بالذكاء الاصطناعي بناء ثقة متوازنة، وتُعزز هذه الثقة الانخراط، ثم يتحول الانخراط إلى استفادة فعلية من التغذية الراجعة.

وتقدم الدراسة تصورًا يركز على المتعلم لفهم التعاون بين الإنسان والذكاء الاصطناعي في تقييم اللغة، وتؤكد أن الإلمام بالذكاء الاصطناعي، لا تطور الأدوات وحده، هو عنصر محوري في جعل التقييم المدعوم به مسؤولاً وذًا قيمة تربوية. كما تناقش الدراسة دلالات النتائج على تصميم التقييم، وإعداد المعلمين، والممارسات الأخلاقية.

الكلمات المفتاحية: التغذية الراجعة المدعومة بالذكاء الاصطناعي؛ الثقة؛ القلق من الذكاء الاصطناعي؛ الإلمام بالذكاء الاصطناعي؛ تقييم اللغة؛ الكتابة باللغة الإنجليزية بوصفها لغةً أجنبية.

Introduction:

Language assessment has always done more than measure proficiency; it shapes what learners attend to, how they revise, and how they come to see themselves as writers. Within computer-assisted language learning (CALL), assessment has steadily absorbed digital tools that deliver feedback, adapt to performance, and generate analytics, making feedback more immediate and more personalised (Chapelle, 2006). The recent diffusion of artificial intelligence (AI), and of large language models (LLMs) in particular, has accelerated this shift. Automated writing evaluation (AWE), automated essay scoring, and LLM-based tools now provide learners with detailed written corrective feedback within seconds, at a scale that no individual teacher could match (Voss, 2024). In writing instruction especially, the practical question is no longer whether machines can generate feedback, but how learners and teachers should work alongside them.

This repositioning turns assessment into a form of human-AI collaboration. Feedback that was once the teacher's province is increasingly produced, or at least drafted, by a system whose reasoning is opaque to the learner, and the learner must decide how far to trust it, how to reconcile it with human advice, and how to act on it (Pack & Maloney, 2024). These are not merely technical decisions. They draw on learners' emotional responses to the technology, their beliefs about its competence, and their capacity to interpret and interrogate what it produces. As several commentators have argued, the value of AI-mediated assessment depends as much on learners' agency and AI literacy as on the tools themselves (Hauck, 2025; Stockwell, 2024).

Despite a rapidly growing literature on AWE effectiveness, comparatively little work has examined the learner-side dispositions that govern whether AI-mediated feedback is taken up at all. Three constructs are especially salient. Trust determines whether learners treat machine feedback as credible enough to act on. Anxiety, well documented in foreign language learning, may be amplified or reshaped when feedback comes from an unfamiliar and non-human source. AI literacy - the knowledge and skills needed to understand, evaluate, and use AI critically - plausibly underpins both, enabling learners to calibrate trust and to manage anxiety rather than be governed by it. How these dispositions combine to shape learners' engagement with, and uptake of, AI-mediated written corrective feedback remains underexplored, particularly in EFL higher-education contexts where such tools are being adopted quickly and unevenly.

This study addresses that gap. It asks how EFL learners' trust in AI feedback, AI anxiety, and AI literacy relate to their engagement with, and uptake of, AI-mediated written corrective feedback, and it uses learners' own accounts to explain the mechanisms behind these relationships. Two commitments shape the framing. First, AI-generated feedback is treated here as a form of formative assessment, not merely as writing support: when a tool flags an error, ranks a draft, or explains a correction, it is making an assessment judgement about the learner's language, and the learner must decide what that judgement is worth. Second, and following from this, trust becomes the hinge of the assessment process, because assessment information only improves learning if the learner acts on it. A sound judgement that the learner distrusts and ignores changes nothing, and a flawed judgement that the learner over-trusts can mislead. Whether AI-mediated assessment works therefore turns on a question of validity-in-use: not only whether the tool's judgements are defensible, but whether the learner has the literacy to weigh them well. (throughout, the term "AI-mediated feedback" is used for the feedback process and "AI feedback" as a shorthand for the same construct; the two are used consistently and interchangeably).

In doing so, the study responds directly to the special issue's call to move beyond simple human-versus-machine comparisons and to examine how AI reshapes the purposes, practices, and outcomes of assessment. Its contribution is both empirical and theoretical. Empirically, it provides a validated model linking learner dispositions to feedback behaviour and a profile analysis showing that these dispositions cluster into recognisable learner types. Theoretically, it proposes a Calibrated Trust Framework for AI-mediated assessment: AI literacy supports calibrated trust and dampens anxiety, calibrated trust drives engagement, and engagement converts into uptake. The framework recasts the familiar problem of when to believe an automated assessment as a matter of learner literacy and calibration rather than of tool accuracy alone, and it places AI literacy at the centre of responsible, learner-centred assessment.

Statement of the Problem:

Although AI-mediated feedback tools are now widely available to English as a foreign language (EFL) learners, their effectiveness cannot be assumed from their technical accuracy alone. The central problem this study addresses is that the value of AI-generated feedback depends on whether learners engage with it and act on it, and this in turn depends on learner dispositions - trust, anxiety, and AI literacy - that remain poorly understood. Learners may distrust and ignore sound feedback, or over-trust and

uncritically adopt flawed feedback; both responses undermine learning. Yet research has concentrated on the tools rather than on the learners who must judge them, leaving a gap in our understanding of how dispositional factors shape feedback engagement and uptake in AI-mediated assessment.

Research Questions:

The study is guided by two research questions:

- 1- How do EFL learners' trust in AI feedback, AI-related anxiety, and AI literacy relate to their engagement with, and uptake of, AI-mediated written corrective feedback?
- 2- How do learners exercise and regulate trust in AI feedback in practice, and what role does AI literacy play in this process?

Objectives of the Study:

Accordingly, the study aims to: (1) model the relationships among AI literacy, trust, anxiety, feedback engagement, and feedback uptake; (2) identify distinct learner profiles based on these dispositions; (3) examine how AI anxiety changes over a period of AI-mediated collaboration and what predicts that change; and (4) characterise, through learners' own accounts, how calibrated trust is enacted and how AI literacy regulates engagement with AI feedback.

Significance of the Study:

This study is significant on three grounds. Theoretically, it advances a learner-centred account of human-AI collaboration in language assessment by proposing and testing a Calibrated Trust Framework, shifting the focus from tool accuracy to learner literacy and calibration. Methodologically, it combines structural equation modelling, profile analysis, and longitudinal and qualitative evidence to provide a multi-faceted account of how learners engage with AI feedback. Practically, the findings inform assessment design, teacher education, and the responsible integration of AI feedback tools, with particular relevance for EFL contexts in which access to individualised teacher feedback is limited and AI tools are increasingly relied upon.

Hypotheses:

To keep the model parsimonious and testable, the study advances three main hypotheses, each integrating several related paths. H1 (Literacy and dispositions): AI literacy positively predicts trust in AI feedback and negatively predicts AI anxiety. H2 (Dispositions and feedback behaviour): trust positively predicts feedback engagement and uptake, whereas AI anxiety negatively predicts engagement and uptake, with engagement contributing to uptake. H3 (Change over time): AI anxiety declines over the

period of AI-mediated collaboration, and higher AI literacy and higher trust each predict a steeper decline.

Literature Review:

AI-mediated feedback as formative assessment in CALL

Automated writing evaluation has a long lineage in CALL, from early essay-scoring engines to today's LLM-based assistants that comment on grammar, lexis, organisation, and content (Voss, 2024; Warschauer & Ware, 2006). Research on tools such as Grammarly, Criterion, and Write & Improve has reported gains in accuracy and in some dimensions of writing quality, alongside high learner satisfaction with the speed and availability of feedback (Koltovskaia, 2020; Ranalli, 2021). The arrival of general-purpose LLMs has widened this scope, allowing learners to request explanations, revisions, and dialogic clarification rather than receiving a fixed score (Barrot, 2023). These developments blur the boundary between learning and assessment: the same tool that evaluates a draft also teaches the learner how to improve it, making feedback formative by design.

Framing these tools as assessment, rather than as convenience features, brings the relevant scholarship into view. Recent work has tested large language models as scorers and feedback givers and has found them capable but uneven, which is exactly why their judgements call for interpretation rather than automatic acceptance (Mizumoto & Eguchi, 2023; Yan et al., 2024). Comparative studies of AI-generated and peer or teacher feedback report that learners weigh sources differently and do not treat machine feedback as authoritative by default (Banihashem et al., 2024). This is where assessment literacy matters. The assessment literacy tradition holds that using assessment information well is itself a competence (Stiggins, 1991), and the related idea of student feedback literacy describes the capacity to make sense of feedback and act on it (Carless & Boud, 2018). Extended to AI, these traditions imply an emerging assessment-specific literacy: knowing what an automated judgement is and is not and deciding accordingly. The present study treats that capacity as central rather than incidental.

Crucially, providing feedback is different from learners using it. The written corrective feedback literature has long distinguished feedback provision from learner engagement and uptake — the cognitive, affective, and behavioural ways in which learners attend to feedback and incorporate it into revision (Ellis, 2010; Han & Hyland, 2015). Machine-generated feedback complicates engagement further, because learners must judge the credibility of an opaque source and decide how to weigh it against human input. Studies of AWE use report that learners frequently accept surface-

level corrections while ignoring or misinterpreting more substantive comments, and that uptake varies widely across individuals (Koltovskaia, 2020; Zhang & Hyland, 2018). This variation motivates attention to the learner dispositions that shape it.

Trust in AI -mediated feedback:

Trust is the willingness to rely on a system under conditions of uncertainty and vulnerability, and in human-automation research it is the central determinant of whether people accept or override machine outputs (Lee & See, 2004). Trust that is well matched to a system's actual reliability - calibrated trust - supports good decisions, whereas over-trust produces uncritical acceptance and under-trust produces disuse. In the assessment context, a learner who over-trusts AI-mediated feedback may adopt erroneous corrections and surrender authorial control, while a learner who under-trusts it may discard useful guidance. Trust is therefore expected to be a strong positive predictor of both engagement and uptake, but its desirability is conditional on its calibration, a point the qualitative phase of this study examines directly. Research on trust in educational AI specifically has begun to show that trust is neither automatic nor uniform: it depends on perceived accuracy, transparency, and the user's understanding of the system, and it can be deliberately developed (Choung et al., 2023; Nazaretsky et al., 2022). This work reinforces the view taken here that trust in AI assessment is a learner disposition open to education, not a fixed trait.

AI anxiety and foreign language learning:

Foreign language anxiety is a well-established affective barrier to learning, associated with avoidance, reduced participation, and poorer performance (Horwitz et al., 1986). A parallel construct, AI anxiety, captures apprehension specifically about interacting with AI systems - fear of misunderstanding the technology, of being judged by it, or of becoming dependent on it (Wang & Wang, 2022). For EFL writers, these two anxieties may compound: feedback that is dense, rapid, and machine-generated can feel overwhelming, particularly for less proficient learners, prompting disengagement. Anxiety is therefore hypothesised to depress engagement and uptake. Because anxiety and trust are conceptually distinct but related - uncertainty about a system can simultaneously lower trust and raise anxiety - both are modelled together rather than in isolation.

AI literacy: a multidimensional construct:

AI literacy refers to the competencies that allow people to understand how AI works, evaluate its outputs critically, use it effectively, and recognise its ethical limits (Ng et al., 2021). In the feedback context, AI literacy should help learners interpret why a tool flags an item, distinguish reliable from unreliable suggestions, and integrate machine feedback with

human advice. It is therefore theorised here as an antecedent that shapes the affective dispositions of trust and anxiety: learners who understand the tools are better placed to extend trust where warranted and to contain anxiety born of unfamiliarity. AI literacy is also expected to support engagement directly, by giving learners strategies for working with feedback. This positions AI literacy not as a peripheral skill but as the regulating capacity at the centre of human-AI collaboration in assessment, consistent with calls to foreground learner agency and literacy in AI-mediated assessment (Hauck, 2025; Stockwell, 2024). Because AI literacy is the pivotal construct in this study, it is worth specifying what it comprises. We adopt the framework of Ng et al. (2021), who synthesise the literature into four dimensions: knowing and understanding AI (how systems such as language models work and where they fail), using and applying AI (operating the tools for a purpose), evaluating and creating with AI (judging outputs critically and adapting them), and AI ethics (recognising fairness, bias, and accountability). Long and Magerko (2020) describe a parallel set of competencies, emphasising that critical evaluation, not mere operation, is the heart of literacy. Mapped onto assessment, these dimensions describe an assessment-specific AI literacy: a learner who understands how a tool reaches a judgement, can appraise that judgement against other evidence, and is alert to its ethical limits is equipped to decide what to accept. It is this evaluative core, rather than fluent button-pressing, that we expect to regulate trust and temper anxiety.

Human-AI collaboration and the Calibrated Trust Framework:

Synthesising these strands, this study frames AI-mediated feedback as a human-in-the-loop process in which the learner remains the decisive agent. AI literacy is modelled as an antecedent of trust and anxiety; trust and anxiety, in turn, shape feedback engagement; and engagement, together with trust and anxiety, predicts feedback uptake. The model thus treats uptake, the behavioural endpoint that matters for learning, as the joint product of what learners know (AI literacy), how they feel (anxiety), what they believe about the tool (trust), and how they act on it (engagement). We name this account the Calibrated Trust Framework for AI-mediated assessment. Its claim is specific: the value of an automated assessment is realised not at the moment the tool produces a judgement but at the moment a learner, equipped with AI literacy, decides how far to trust and act on it. AI literacy is the antecedent that makes trust calibrated rather than blind and keeps anxiety from becoming avoidance; calibrated trust is the proximal driver of engagement; and engagement carries through to uptake. The following hypotheses operationalise this framework: The eight structural paths below (H1-H8) and three longitudinal paths (H9-H11) elaborate the three main

hypotheses stated earlier (H1–H3): they are retained here as the specific, separately estimated paths that the structural and longitudinal models test.

H1: AI literacy positively predicts trust in AI - mediated feedback.

H2: AI literacy negatively predicts AI anxiety.

H3: Trust positively predicts feedback engagement.

H4: Trust positively predicts feedback uptake.

H5: AI anxiety negatively predicts feedback engagement.

H6: AI anxiety negatively predicts feedback uptake.

H7: Feedback engagement positively predicts feedback uptake.

H8: AI literacy positively predicts feedback engagement.

Because the Calibrated Trust Framework treats AI literacy and trust as resources that should help learners settle into working with AI-mediated feedback, it also makes predictions about change over time. As learners gain experience within the human-AI collaboration, their anxiety should fall, and it should fall faster for those who understand the tools and trust them. Three longitudinal hypotheses capture this:

H9: AI anxiety declines significantly over the period of AI-mediated collaboration.

H10: Higher AI literacy predicts a steeper decline in AI anxiety (a more negative growth slope).

H11: Higher trust predicts a steeper decline in AI anxiety (a more negative growth slope).

Two research questions framed the qualitative phase, which was designed to explain the survey results rather than merely corroborate them:

- 1-How do EFL learners describe trusting, and acting on, AI-mediated writing feedback?
- 2-What role does AI literacy play in how learners manage trust and anxiety when human and machine feedback are both available?

AI feedback in Arab and Gulf EFL contexts:

While much of the research on AI-mediated feedback has been conducted in Western and East Asian settings, a growing body of work situates these questions in Arab and Gulf higher education, where this study is located. In Saudi Arabia specifically, Alsofyani and Barzanji (2025) compared ChatGPT-generated and teacher feedback on tertiary EFL learners' writing and found AI feedback to be a viable complement to teacher feedback, particularly given large class sizes that limit individualised teacher response. Elmotri et al. (2025) examined automated written corrective feedback at a Saudi university and reported that learners found such systems usable and useful, while preferring specific over generic feedback — a finding that foregrounds learner judgement of feedback quality rather than tool capability alone. Khasawneh (2025), in a Saudi

sample, found that AI-generated corrective feedback together with self-regulation accounted for a substantial share of variance in language proficiency, underscoring the role of learner-side regulatory resources that the present study conceptualises as AI literacy and calibrated trust.

These regional studies establish that AI feedback tools are already embedded in Gulf EFL writing instruction and that learner dispositions, not merely tool accuracy, shape their effect. They also reveal a gap the present study addresses: regional work has largely documented perceptions and proficiency outcomes, but has not modelled how trust, anxiety, and AI literacy jointly drive engagement and uptake, nor how these dispositions cluster into learner profiles. Situating the present study in this context therefore strengthens its relevance to Arab and Gulf universities while addressing a substantive, not merely geographic, gap in the literature.

Method:

Design:

An explanatory sequential mixed-methods design with a longitudinal quantitative strand was used (Creswell & Plano Clark, 2018). Learners were surveyed at three points across a semester of working with AI-mediated feedback, which allowed both a cross-sectional test of the relationships among dispositions and a longitudinal analysis of how AI anxiety changed over the period of collaboration. A qualitative interview phase with a purposive subsample then explained and contextualised the statistical patterns. The relational (structural) model was estimated on the final-wave data, when learners had the most experience to draw on, while the analysis of change in anxiety used all three waves. This design fits the study's aim of estimating relationships among dispositions, tracking how anxiety shifts as learners gain experience, and understanding the reasoning behind both, and it matches the special issue's interest in empirically grounded accounts that move beyond surface comparisons.

Participants and context:

Participants were 300 undergraduate EFL learners recruited from several Saudi universities. Participants were enrolled in undergraduate English language courses offered across participating Saudi universities and where AWE and LLM-based tools are increasingly used in writing courses. Participants were recruited through writing instructors and took part voluntarily. All had used at least one AI-based writing tool (most commonly Grammarly, Write & Improve, or an LLM assistant) during the semester. Of an initial (327) learners at Wave 1, complete data across all three waves were available for (300) (an 8.3% attrition rate); analyses use this retained sample, and attrition analyses showed no significant baseline differences

between completers and non-completers on the study variables. The interview subsample comprised (16) learners selected to vary in proficiency, prior tool use, and survey responses regarding trust and anxiety, so as to represent a range of experiences. Sample characteristics are reported in Table (1).

Table (1)

Participant Characteristics (N = 300)

Characteristic	Category	n (%)
Gender	Female	182 (60.7%)
	Male	118 (39.3%)
Year of study	First year	96 (32.0%)
	Second year	84 (28.0%)
	Third year	72 (24.0%)
	Fourth year	48 (16.0%)
Proficiency (CEFR)	B1	121 (40.3%)
	B2	138 (46.0%)
	C1	41 (13.7%)
Prior use of AWE tools	Regular (weekly)	94 (31.3%)
	Occasional	146 (48.7%)
	None/rare	60 (20.0%)

Note. CEFR = Common European Framework of Reference; AWE = automated writing evaluation. Proficiency was determined using placement scores provided by the participating universities.

Instruments:

The survey comprised five multi-item reflective scales adapted from validated instruments and contextualised to AI-mediated writing feedback. AI literacy was measured with six items adapted from Ng et al. (2021) and Wang and Wang (2022). Trust in AI feedback (five items) was adapted from human-automation trust measures (Lee & See, 2004) and worded for the feedback context. AI anxiety (five items) drew on foreign language anxiety (Horwitz et al., 1986) and AI anxiety measures (Wang & Wang, 2022). Feedback engagement (five items) captured cognitive and behavioural attention to feedback (Ellis, 2010; Han & Hyland, 2015), and feedback uptake (four items) captured the extent to which learners incorporated feedback into revision. The AI anxiety scale was administered at all three waves to enable longitudinal modelling of change. AI literacy and trust were

measured at Wave (1) and again at Wave (3), allowing their Wave (1) scores to serve as baseline predictors in the latent growth model. Feedback engagement and feedback uptake were measured at the final wave, after participants had accumulated sufficient experience using AI-mediated writing feedback. The instrument was reviewed by two applied-linguistics experts for content validity and piloted with (28) learners who did not take part in the main study; minor wording revisions followed. The full survey instrument, including all items for the five scales, is provided as an appendix (see Appendix A/ Supplementary Materials) to support transparency and replication.

The semi-structured interview protocol invited learners to describe a recent episode of receiving AI feedback on their writing, how they decided what to act on, what they did when AI and teacher feedback diverged, and how confident they felt working with the tools. Probes elicited concrete examples rather than general opinions. Interviews lasted 30-45 minutes, were conducted in the participants' stronger language where helpful, audio-recorded with consent, and transcribed verbatim.

Procedure:

Data were collected over a single academic semester across the participating universities during which participants used AI-mediated feedback as part of their regular writing instruction. The AI anxiety survey was administered online at three points: Wave (1) in the first two weeks (baseline), Wave (2) at mid-semester, and Wave (3) in the final two weeks, where participants completed the second administration of the AI literacy and trust scales together with the feedback engagement and feedback uptake measures. Interviews followed in the two weeks after Wave (3). Responses were linked across waves with an anonymous code, and no course credit was contingent on participation.

Data analysis:

The hypothesised model was tested using partial least squares structural equation modelling (PLS-SEM) in SmartPLS, chosen for its suitability with predictive, theory-building models and its tolerance of non-normal data (Hair et al., 2022). The measurement model was assessed using indicator loadings, composite reliability (CR), and average variance extracted (AVE), and discriminant validity was assessed using the heterotrait-monotrait ratio (HTMT). The structural model was evaluated through path coefficients and their significance (bootstrapping with 5,000 resamples), the coefficient of determination (R^2), the standardised root mean square residual (SRMR) as a measure of approximate fit, collinearity among predictors (variance inflation factor, VIF), effect sizes (Cohen's f^2), and predictive relevance (Stone-

Geisser Q^2 via blindfolding). Reporting these additional diagnostics follows current guidance for PLS-SEM (Hair et al., 2022).

Change in AI anxiety over the three waves was examined in two complementary ways. First, a mixed-design repeated-measures ANOVA tested the within-subjects effect of time (three waves) and its interaction with a between-subjects factor formed by splitting AI literacy into low, medium, and high groups; Mauchly's test checked the sphericity assumption, with the Greenhouse-Geisser correction applied where it was violated, and Bonferroni-corrected pairwise comparisons followed significant effects. Second, a latent growth curve model (LGCM) was fitted to the three anxiety measures to estimate each learner's trajectory as a latent intercept (initial anxiety) and a latent slope (rate of change). An unconditional model established whether anxiety declined on average and whether learners differed in their starting points and rates; a conditional model then added Wave 1 AI literacy and trust as predictors of the intercept and slope, to test whether these dispositions shaped how far and how fast anxiety fell. Model fit was judged with the chi-square test, CFI, TLI, RMSEA, and SRMR.

To test whether learners formed distinct dispositional types rather than varying along a single continuum, we ran a cluster analysis on the five standardised construct scores. The number of clusters was decided before interpretation: Ward's hierarchical method and inspection of the dendrogram and agglomeration schedule suggested a small number of groups, and we compared two-, three-, and four-cluster solutions using the average silhouette width, choosing the solution with the best-separated and most interpretable structure. Final membership was estimated with k-means, and cluster centroids are reported as standardised means so that each profile can be read against the sample average.

Interview data were analysed using reflexive thematic analysis (Braun & Clarke, 2006): transcripts were read repeatedly, coded inductively, and organised into themes that were reviewed against the data and against the quantitative findings. To strengthen trustworthiness, coding was discussed among the research team, an audit trail was kept, and illustrative extracts are reported below.

Ethical considerations:

The study followed institutional ethical guidelines. Participation was voluntary and could be withdrawn at any point without penalty; the study posed minimal risk, involving reflection on ordinary coursework rather than any sensitive intervention, and its benefits - informing more supportive feedback practice - were judged to outweigh that minimal risk. Informed

consent was obtained in writing, data were anonymised and stored securely, and pseudonyms (L1-L16) are used for interview participants throughout.

The analytical sequence is deliberately layered rather than redundant, with each component answering a distinct question that the others cannot. The PLS-SEM model tests the hypothesised dispositional pathways to engagement and uptake (RQ1); the cluster analysis asks whether these dispositions combine into recognisable learner profiles rather than operating only as variable-level effects; the repeated-measures ANOVA and latent growth curve model address whether and for whom AI anxiety changes over time, which the cross-sectional structural model cannot establish; and the qualitative phase explains the mechanisms behind these patterns (RQ2). PLS-SEM was selected partly because it is appropriate for the present sample size and for predictive, theory-building models, and the longitudinal and qualitative components were kept deliberately focused to avoid overextending the data. The aim is convergence across complementary lenses, not an accumulation of analyses for their own sake.

Results:

Measurement model:

All indicator loadings exceeded the (.70) threshold, and composite reliability for every construct was above (.80), indicating satisfactory internal consistency. Average variance extracted exceeded, (.50) for all constructs, supporting convergent validity (Table 2). Discriminant validity was supported, as all HTMT ratios were below the conservative (.85) criterion (Table 3). The measurement model was therefore judged adequate for structural assessment.

Table (2)

Measurement Model: Reliability and Convergent Validity

Construct	Items	Loadings	CR	AVE
AI literacy (AIL)	6	.84–.81	.89	.58
Trust in AI feedback (TRU)	5	.86–.78	.90	.64
AI anxiety (ANX)	5	.83–.74	.87	.58
Feedback engagement (ENG)	5	.82–.76	.88	.59
Feedback uptake (UPT)	4	.85–.79	.89	.67

Note. CR = composite reliability; AVE = average variance extracted. Loadings are reported as a range across items within each construct.

Table (3)
Discriminant Validity (HTMT Ratios)

Construct	1	2	3	4	5
1. AIL	—				
2. TRU	.54	—			
3. ANX	.48	.51	—		
4. ENG	.57	.62	.45	—	
5. UPT	.50	.66	.49	.69	—

Note. AIL = AI literacy; TRU = trust in AI feedback; ANX = AI anxiety; ENG = feedback engagement; UPT = feedback uptake. All ratios fall below the .85 threshold.

Structural model and hypothesis tests:

The structural model explained substantial variance in the endogenous constructs: R^2 was .48 for feedback uptake, (.43) for feedback engagement, (.21) for trust, and (.15) for anxiety. All eight hypothesised paths were statistically significant (Table 4). AI literacy positively predicted trust ($H1$; $\beta = .46$) and negatively predicted anxiety ($H2$; $\beta = -.39$), supporting its theorised role as a regulating antecedent. Trust was the strongest positive predictor in the model, shaping both engagement ($H3$; $\beta = .41$) and uptake ($H4$; $\beta = .44$). Anxiety exerted smaller but significant negative effects on engagement ($H5$; $\beta = -.28$) and uptake ($H6$; $\beta = -.19$). Engagement positively predicted uptake ($H7$; $\beta = .37$), and AI literacy retained a modest direct effect on engagement beyond its indirect routes ($H8$; $\beta = .18$). The model showed acceptable approximate fit ($SRMR = .046$, below the .08 threshold). Taken together, it indicates that trust is the proximal driver of feedback behaviour, that anxiety dampens it, and that AI literacy operates largely by shaping these two dispositions.

Table (4)
Structural Model: Path Coefficients and Hypothesis Tests

Path (hypothesis)	β	t	p	Decision
H1: AIL $\rightarrow$ TRU	.46	7.84	<.001	Supported
H2: AIL $\rightarrow$ ANX	-.39	6.12	<.001	Supported
H3: TRU $\rightarrow$ ENG	.41	6.77	<.001	Supported
H4: TRU $\rightarrow$ UPT	.44	7.05	<.001	Supported
H5: ANX $\rightarrow$ ENG	-.28	4.36	<.001	Supported
H6: ANX $\rightarrow$ UPT	-.19	3.01	.003	Supported
H7: ENG $\rightarrow$ UPT	.37	5.62	<.001	Supported
H8: AIL $\rightarrow$ ENG	.18	2.71	.007	Supported

Note. β = standardised path coefficient; t and p from bootstrapping with (5,000) resamples. All paths significant at ($p < .01$).

Collinearity, effect sizes, and predictive relevance supported the model (Table 5). All inner-model VIF values were well below the conservative threshold of (3.3), indicating no collinearity problems. Effect sizes (f^2) ranged from small to medium, with the trust paths and the engagement-to-uptake path carrying the largest effects, consistent with the central role the framework assigns to trust. Blindfolding produced positive Q^2 values for every endogenous construct (trust = .13, anxiety = .09, engagement = .26, uptake = .31), confirming the model's predictive relevance for the outcomes that matter, especially uptake.

Table (5)

Collinearity (VIF), Effect Sizes (f^2), and Effect-Size Interpretation

Structural path	VIF	f^2	Effect size
AI literacy → Trust	1.00	.27	Medium
AI literacy → Anxiety	1.00	.18	Medium
AI literacy → Engagement	1.42	.05	Small
Trust → Engagement	1.51	.21	Medium
Anxiety → Engagement	1.33	.11	Small-medium
Trust → Uptake	1.66	.24	Medium
Anxiety → Uptake	1.38	.05	Small
Engagement → Uptake	1.74	.16	Medium

Note. VIF = variance inflation factor (inner model); f^2 benchmarks (Cohen): (.02) small, (.15) medium, (.35) large. Q^2 (blindfolding) was positive for all endogenous constructs (trust .13, anxiety .09, engagement .26, uptake .31), and model SRMR = (.046).

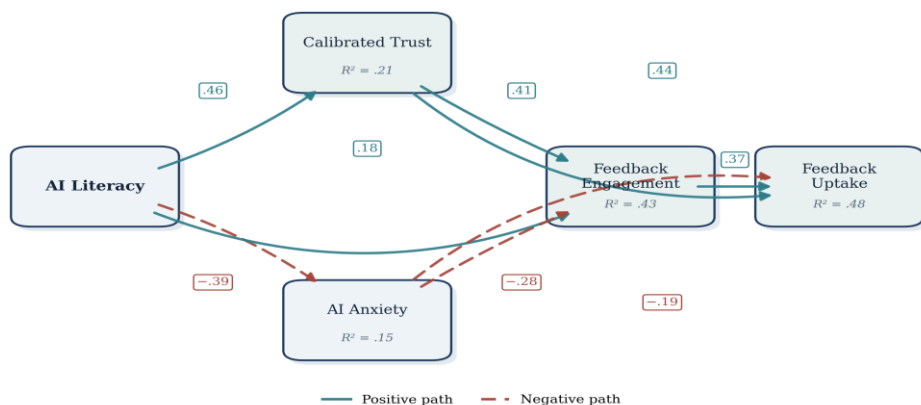


Figure (1) Structural model with standardised path coefficients. Solid lines are positive paths, and dashed lines are negative paths; R^2 values appear inside the endogenous constructs.

Figure (1) brings these results together. Reading it from left to right, AI literacy feeds into how much learners trust the tool and how anxious they feel about it, and those two dispositions then carry through to engagement and, finally, to uptake. Trust does most of the work; anxiety pulls in the opposite direction but with a lighter hand.

Learner profiles: cluster analysis:

The path model tells us how the dispositions hang together on average, but it does not tell us whether learners actually cluster into recognisable types. They did. Ward's method pointed to a three-group solution, and the k-means clusters that followed were easy to read against the five construct scores (Figure 2). We label them Confident Collaborators, Anxious Verifiers, and Detached Low-Literacy Users.

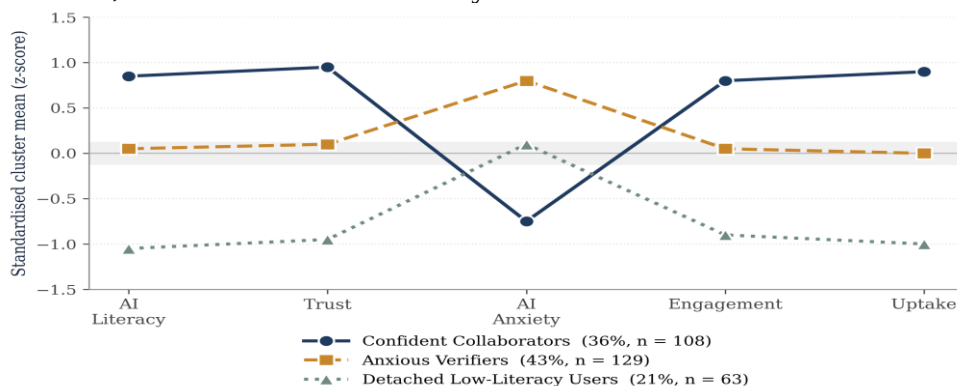


Figure (2) Three learner profiles from k-means clustering, shown as standardised means (z-scores) across the five constructs. Values above zero are above the sample average; values below zero are below it.

The first group, the Confident Collaborators (36%), scored high on AI literacy, trust, engagement, and uptake, and low on anxiety. These are the learners the model would predict do best with AI feedback: they understand the tool, they trust it enough to act on it, and they are not thrown by it. The largest group, the Anxious Verifiers (43%), sat near the average on literacy, trust, engagement, and uptake but well above it on anxiety. They use the feedback and they take some of it up, but they do so warily, and the affective cost is visible in their scores. The third group, the Detached Low-Literacy Users (21%), scored below average almost everywhere, with literacy, trust, engagement, and uptake all low; their anxiety was only mildly elevated, which fits the picture of learners who step back from the tool rather than wrestle with it.

The profiles map neatly onto the path model. Where AI literacy and trust are high and anxiety is low, uptake follows; where literacy is low, the whole chain sags. The middle group is the most instructive, because it shows that anxiety and reasonable uptake can coexist: these learners do not abandon the feedback, but they pay for using it in worry, and they are the clearest candidates for support.

The three-cluster solution was preferred on both statistical and interpretive grounds. The average silhouette width was highest for three clusters (.52), compared with (.45) for two and (.41) for four, indicating that three groups were the best separated; the three-group structure was also the most readable, with each profile telling a coherent story across the five constructs. Standardised centroids and group sizes are reported in Table (6).

Table (6)

Cluster Centroids (Standardised Means) and Group Sizes

Profile	n (%)	AIL	TRU	ANX	ENG	UPT
Confident Collaborators	108 (36%)	0.85	0.95	-0.75	0.80	0.90
Anxious Verifiers	129 (43%)	0.05	0.10	0.80	0.05	0.00
Detached Low-Literacy Users	63 (21%)	-1.05	-0.95	0.10	-0.90	-1.00

Note. Centroids are standardised means (z-scores); values above zero exceed the sample average. AIL = AI literacy; TRU = trust; ANX = anxiety; ENG = engagement; UPT = uptake. Average silhouette width = (.52) for the three-cluster solution.

To make the profiles useful for teaching and assessment, Table (7) characterises each group in plain terms by its three defining dispositions, AI literacy, trust, and anxiety, together with the feedback behaviour these tend to produce and the support each group is likely to need. The three profiles trace the logic of the

Calibrated Trust Framework. The Confident Collaborators combine high AI literacy with high, calibrated trust and low anxiety, and it is this combination that lets them engage closely and take up feedback while still exercising judgement. The Anxious Verifiers sit in the middle on literacy and trust but carry high anxiety; they do the work and benefit from it, yet the effort costs them, which marks them as the group most in need of support with feedback load and confidence. The Detached Low-Literacy Users are low on literacy and trust and, rather than being highly anxious, are largely disengaged; for them the priority is building the basic literacy that would let trust, and therefore use, develop at all.

Table (7)

Learner Profiles by AI Literacy, Trust, and Anxiety

Profile (n, %)	AI literacy	Trust	Anxiety	Typical feedback behaviour	Pedagogical and assessment need
Confident Collaborators (n = 108, 36%)	High	High, calibrated	Low	Engage deeply; verify suggestions, then take up feedback selectively while keeping authorial control	Extension and self-evaluation tasks; safeguards against over-trust; can model verification strategies for peers
Anxious Verifiers (n = 129, 43%)	Moderate	Moderate	High	Use feedback but warily; double-check often; progress real but slowed by worry and feedback overload	Manage feedback load and pacing; confidence-building; scaffolds that convert effortful checking into uptake
Detached Low-Literacy Users (n = 63, 21%)	Low	Low	Mild (disengaged)	Limited engagement and low uptake; tend to step back from the tool rather than work through it	Foundational AI literacy; guided, low-pressure introduction; reduce entry barriers before independent use

Note. Levels (high, moderate, low) describe each cluster relative to the sample, based on the standardised centroids in Table (6). The profiles are group-level types, not individual diagnoses.

The profiles also align with the change observed over time. The high-literacy, high-trust Confident Collaborators were the group whose anxiety fell fastest across the three waves, whereas the Detached Low-Literacy

Users, low on both literacy and trust, showed the flattest decline. Read together with the growth model, the profiles suggest that the learners best placed to benefit from AI-mediated assessment are not simply those who feel calm at the outset, but those whose literacy and trust let their anxiety subside as the collaboration develops, and that the low-literacy group needs deliberate support if it is not to be left behind.

Change in AI anxiety over time:

Because anxiety was measured at three points, we could ask not only how it related to the other constructs but whether it changed as learners settled into working with AI feedback. It did. Across the sample, mean AI anxiety fell from (3.33) at Wave 1 to (2.84) at Wave (2) and (2.39) at Wave (3). The decline, and the way it differed by AI literacy, is shown in Figure (3).

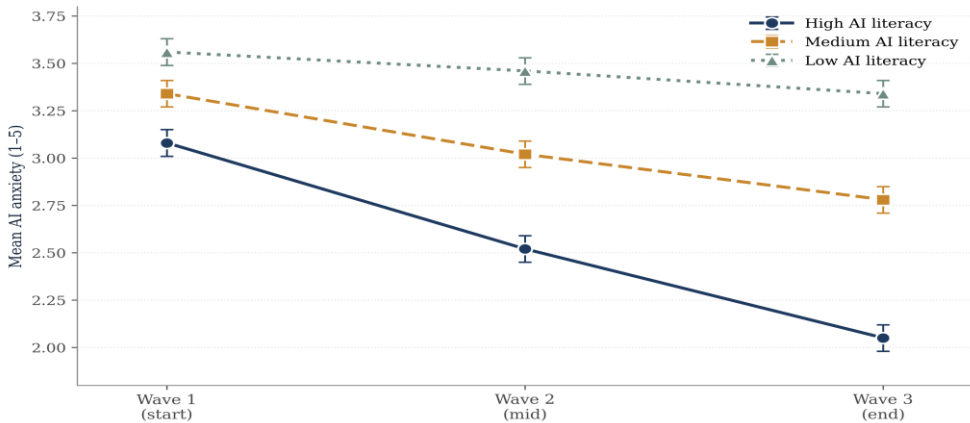


Figure (3) Mean AI anxiety across the three waves by AI literacy group (error bars show ± 1 SE). Anxiety declined in all groups and fell most steeply among high-literacy learners.

A mixed-design repeated-measures ANOVA confirmed the pattern (Table 8). Mauchly's test indicated that sphericity was violated ($W = .91$, $p = .004$), so Greenhouse-Geisser-corrected tests are reported. The within-subjects effect of time was significant and substantial, with anxiety falling across the three waves ($F(1.82, 540.4) = 66.50$, $p < .001$, partial $\eta^2 = .18$). Bonferroni-corrected comparisons showed each wave lower than the one before (Wave 1 > Wave 2 > Wave 3, all $p < .001$). The time-by-literacy interaction was also significant ($F(3.64, 540.4) = 10.20$, $p < .001$, partial $\eta^2 = .06$): higher-literacy learners started lower and dropped faster, while the lowest-literacy group barely moved. The between-subjects effect of literacy was significant as well ($F(2, 297) = 19.30$, $p < .001$, partial $\eta^2 = .12$).

Table (8)

Repeated-Measures ANOVA for AI Anxiety Across Three Waves

Effect	df	F	Partial η^2
Time (within)	1.82, 540.4	66.50***	.183
Time $\times$ AI literacy group	3.64, 540.4	10.20***	.064
AI literacy group (between)	2, 297	19.30***	.115

Note. Greenhouse-Geisser-corrected degrees of freedom reported (Mauchly's $W = .91$, $p = .004$). The between-subjects factor is AI literacy group (low, medium, high). (***) $p < .001$.

The latent growth curve model put numbers on the individual trajectories (Table 9). The unconditional linear model fitted well ($\chi^2(1) = 1.84$, $p = .175$; CFI = .995; TLI = .986; RMSEA = .042, 90% CI [.000, .118]; SRMR = .021). Its mean intercept was 3.33 ($p < .001$), and its mean slope was -0.49 per wave ($p < .001$), so anxiety reliably decreased over the semester. Both the intercept variance (.39, $p < .001$) and the slope variance (.07, $p = .012$) were significant, meaning learners differed both in where they started and in how quickly their anxiety fell, which is the variation the framework predicts AI literacy and trust should explain.

Adding Wave 1 AI literacy and trust as predictors confirmed that they did. Higher AI literacy predicted lower initial anxiety ($\beta = -.33$, $p < .001$) and a steeper decline (a more negative slope; $\beta = -.28$, $p = .002$). Higher trust likewise predicted lower initial anxiety ($\beta = -.24$, $p = .003$) and a steeper decline ($\beta = -.20$, $p = .013$). Together the predictors explained 17–19% of the variance in the growth factors, and the conditional model retained acceptable fit (CFI = .972; RMSEA = .051; SRMR = .039). In plain terms, learners who understood the tools and trusted them did not just feel less anxious at the outset; their anxiety also came down faster as the collaboration went on.

Table (9)

Latent Growth Curve Model of AI Anxiety: Unconditional Estimates and Predictors

Parameter	Estimate	β (predictor)	R ²
Mean intercept (initial anxiety)	3.33***	—	—
Mean slope (change per wave)	−0.49***	—	—
Intercept variance	.39***	—	—
Slope variance	.07*	—	—
AI literacy → intercept	—	−.33***	—
AI literacy → slope	—	−.28**	—
Trust → intercept	—	−.24**	—
Trust → slope	—	−.20*	—
Variance explained (R ²)	—	—	.17–.19

Note. Intercept = initial anxiety; slope = change per wave (negative = decline). Predictor columns report standardised β from the conditional model, with Wave 1 AI literacy and trust predicting the growth factors. (* $p < .05$), (** $p < .01$), (***) $p < .001$).

Read alongside the cross-sectional model, these results extend the Calibrated Trust Framework into time: AI literacy and trust are not only associated with lower anxiety at a single moment but actively shape how quickly anxiety subsides as learners work with AI feedback. The human-AI collaboration itself appears to reduce anxiety, and it does so fastest for the learners best equipped to make sense of the tools.

Qualitative findings:

Five themes were developed from the interviews; their prevalence across the (16) participants and a brief description are summarised in Table (10). The themes elaborate the survey results by showing how learners reasoned about trusting and using AI feedback.

Table (10)

Themes from Learner Interviews (n = 16)

Theme	Participants	Description
Calibrated trust through verification	11 of 16	Learners cross-checked AI feedback against teachers, peers, or reference sources before accepting it.
Anxiety from feedback overload	9 of 16	Dense, simultaneous corrections produced fatigue and avoidance, especially among lower-proficiency learners.
AI literacy as a regulating resource	10 of 16	Understanding how the tools work helped learners interpret, question, and selectively adopt feedback.
Negotiating authority when sources diverge	8 of 16	Learners deferred to teachers when AI and human feedback conflicted, treating AI as a first-pass advisor.
Shifting ownership of writing	7 of 16	Concerns about over-reliance and loss of voice prompted deliberate limits on AI use.

Note. Participant counts indicate how many of the (16) interviewees expressed each theme; pseudonyms L1–L (16) identify participants.

Calibrated trust through verification (RQ1):

Most learners (11 of 16) described trusting AI feedback only after checking it, rather than accepting it wholesale. They treated the tool as a capable but fallible first reader whose suggestions had to earn acceptance. As L (4) explained, the machine was useful precisely because it was provisional:

"I don't just click accept. If it changes my sentence, I read it and ask myself whether it still says what I meant. Sometimes it's right, sometimes it changes my meaning, so I check with my teacher or a dictionary".

This pattern mirrors the strong but conditional role of trust in the survey model. Learners who trusted the tool still retained authorial judgement, and several described becoming better at deciding what to accept as the semester progressed - an account of trust becoming better calibrated with experience.

Anxiety from feedback overload (RQ1):

Nine of the sixteen learners described feeling overwhelmed when the tool returned many corrections at once, an experience that sometimes led them to disengage. The affective load was heaviest for lower-proficiency learners, consistent with the negative path from anxiety to engagement. As L (9) put it:

"When I see twenty red marks, I feel like my writing is terrible and I don't want to look at it. It is too much at one time, so I just close it and come back later, or sometimes I don't".

Notably, learners who reported strategies for working through feedback in manageable steps described less of this avoidance, foreshadowing the regulating role of AI literacy.

AI literacy as a regulating resource (RQ2):

Ten of the sixteen learners connected their confidence with feedback to some understanding of how the tools worked. Those who grasped that the system pattern-matches rather than understands meaning were more willing to question its suggestions and less unsettled by them. L13 described this regulating function directly:

"Once I understood it is guessing the most likely correction, not reading my mind, I stopped being afraid of it. I use it, but I know it can be wrong, so I decide".

This account explains why, in the survey model, AI literacy raised trust and lowered anxiety: understanding the tool gave learners a basis for extending trust where it was warranted and for containing the apprehension that came from treating the system as an inscrutable authority.

Negotiating authority when sources diverge (RQ1, RQ2):

When AI and teacher feedback conflicted, eight of the sixteen learners deferred to the teacher, positioning the AI as an advisor rather than an arbiter. Several, however, described using the divergence productively - taking it as a prompt to investigate rather than simply siding with one source. This learner-as-decider stance is the practical expression of the human-in-the-loop framing and suggests that divergence between human and machine feedback can be pedagogically useful when learners have the literacy to exploit it.

Shifting ownership of writing (RQ1):

Finally, seven learners voiced concern about over-reliance and the loss of their own voice, and some deliberately limited their use of AI feedback to protect their authorial ownership. This concern, expressed most often by more proficient and more AI-literate learners, indicates that calibrated trust includes knowing when not to use the tool - a dimension of responsible use that purely accuracy-focused evaluations of AWE tend to miss.

These findings carry particular weight for developing-country and Arab university contexts, including the Gulf setting of the present study, where class sizes are often large and individualised teacher feedback is correspondingly scarce (see Section 2.6). In such settings, AI-mediated feedback is not a supplement to abundant teacher attention but frequently the primary source of rapid, detailed feedback available to learners. The

verification practices, literacy-driven regulation of anxiety, and deliberate boundary-setting around AI use documented here therefore describe more than a general disposition: they describe the specific competencies EFL learners in under-resourced and rapidly AI-adopting contexts need in order to benefit from AI-mediated feedback safely, rather than either rejecting it or depending on it uncritically in the absence of sufficient human feedback to cross-check against.

Discussion:

This study examined how EFL learners' trust, anxiety, and AI literacy relate to their engagement with, and uptake of, AI-mediated written corrective feedback, and used learners' accounts to explain those relationships. Three findings stand out. First, trust was the proximal driver of feedback behaviour, predicting both engagement and uptake more strongly than any other construct. This aligns with human-automation research positioning trust as the gatekeeper of reliance (Lee & See, 2004), and it confirms that, in assessment, the decisive question is not whether the tool is accurate but whether the learner is willing to act on it. The qualitative data sharpen this point: the trust that mattered was calibrated trust, extended after verification, not blind acceptance. Learners treated AI as a capable first reader whose suggestions had to be checked, which is precisely the disposition that responsible AI-mediated assessment requires.

Second, anxiety depressed engagement and uptake, but less powerfully than trust raised them. The interviews located this anxiety in feedback overload rather than in the technology as such and showed it falling most heavily on lower-proficiency learners. The cluster analysis made the same point in a different way. The largest group, the Anxious Verifiers, used the feedback and took some of it up while still scoring high on anxiety, which tells us that anxiety and engagement are not simple opposites; a learner can push through worry and still benefit, but at a cost worth taking seriously. This extends the foreign language anxiety tradition (Horwitz et al., 1986) into the AI-mediated context and suggests a design implication: the volume and sequencing of machine feedback, not only its content, but shape also whether learners can engage with it. Tools that surface corrections gradually, or that let learners control the flow, may reduce the affective barrier that dense automated feedback creates.

The longitudinal results add a hopeful qualification to this picture. Anxiety was not fixed: it fell markedly over the semester of AI-mediated collaboration, and the growth model showed that the fall was steeper for learners who began with more AI literacy and more trust. This matters for the framework in two ways. It suggests that working with AI feedback,

rather than simply being exposed to it, can itself reduce the apprehension that surrounds it, and it shows that the same dispositions which drive uptake also govern how quickly that apprehension subsides. The learners least equipped to interpret the tools, by contrast, stayed anxious, which is a warning that time alone does not level the field. Building AI literacy early looks less like a nicety than like the lever that lets anxiety come down for everyone, not only for those who arrive confident.

Third, and most consequentially for the special issue's concerns, AI literacy functioned as the regulating capacity at the centre of the model. It raised trust, lowered anxiety, and retained a direct link to engagement, and the interviews explained why: understanding how the tools work gave learners a basis for calibrating trust and for containing the apprehension that comes from treating an opaque system as an authority. This positions AI literacy not as an optional add-on but as the condition under which human-AI collaboration in assessment becomes productive rather than either over-reliant or avoidant (Hauck, 2025; Ng et al., 2021; Stockwell, 2024). It also reframes the much-discussed problem of divergence between human and machine assessment: when learners have the literacy to interpret it, divergence becomes an occasion for learning rather than a crisis of authority.

Together these findings support a learner-centred, human-in-the-loop account of AI-mediated assessment. The behavioural endpoint that matters for learning - uptake - was the joint product of what learners knew, how they felt, what they believed about the tool, and how they acted on it. This integrative picture answers the special issue's call to move beyond human-versus-machine comparisons and to examine how AI reshapes the purposes and practices of assessment. The machine does not replace the learner's judgement; it relocates that judgement to a new task, deciding what machine feedback is worth and what to do about it, for which AI literacy is the essential preparation.

This is the substance of the Calibrated Trust Framework we propose. Its theoretical claim is that the validity of an AI-mediated assessment is not settled by the tool's accuracy alone; it is completed, or undone, in the learner's decision about what to act on. AI literacy supports calibrated trust and restrains anxiety, calibrated trust drives engagement, and engagement converts into uptake. Framing the contribution this way connects three literatures that are usually kept apart, namely trust in automation, language-learner affect, and AI literacy, and it gives assessment designers and teachers a specific target: not maximal trust, but well-calibrated trust, and the literacy that produces it. It also reframes validity in AI-mediated assessment as partly a property of the user, an argument with clear

consequences for how such assessment should be designed, taught, and judged.

Implications:

Pedagogical implications:

If trust drives uptake and AI literacy calibrates trust, then writing programmes should teach learners not only to use AWE and LLM tools but to interrogate them. Brief, explicit instruction in how the tools generate feedback, why they err, and how to verify suggestions is likely to yield better feedback uptake than simply providing access. Teachers can model calibrated trust by working through divergent human and machine feedback with learners, treating disagreement as a thinking prompt. Staging feedback so that learners are not confronted with all corrections at once may also reduce the overload-driven anxiety the interviews identified.

Implications for assessment design:

The findings suggest that the design of AI-mediated assessment should attend to affective load and to learner control, not only to accuracy. Features that let learners request feedback incrementally, that explain the basis for a suggestion, and that flag uncertainty in the system's own output would support calibrated trust and reduce anxiety. Designing for the human-in-the-loop - keeping the learner as the decisive agent rather than positioning the tool as an arbiter - is consistent with both the evidence here and the ethical aspirations of responsible AI in assessment.

Ethical implications:

Learners' concern about over-reliance and loss of voice points to an ethical dimension that accuracy metrics overlook. Responsible AI-mediated assessment should preserve learner agency and authorial ownership and should be attentive to equity: because anxiety fell most heavily on lower-proficiency learners, tools and practices that ignore affective load risk widening rather than narrowing gaps. AI literacy is therefore not only a competence but an equity measure, distributing the capacity to use these tools well rather than leaving it to those who already arrive confident.

These implications carry particular weight in developing-country and Arab university contexts, including the Gulf setting of this study, where access to individualised teacher feedback is often constrained by large class sizes and heavy teaching loads. In such settings, AI feedback tools are not a marginal supplement but an increasingly central source of written corrective feedback, which makes learners' capacity to judge that feedback - their calibrated trust and AI literacy - especially consequential. The findings suggest that, where teacher feedback is scarce, investing in learners' AI

literacy may be among the most efficient ways to make AI-mediated feedback pedagogically productive rather than anxiety-inducing.

Limitations and Future Research:

Several limitations qualify these findings. The relational model is cross-sectional, and all measures are self-reported, so the structural paths capture associations rather than causal effects, and behavioural uptake was measured by learner report rather than by direct observation of revisions. The anxiety analyses are longitudinal across three waves, which strengthens the case that anxiety genuinely declined, but without a no-treatment control group the decline cannot be attributed to the AI-mediated collaboration with certainty rather than to maturation or other co-occurring instruction. The study was conducted at a single institution with one population of EFL learners, which bounds transferability. The interview sample, while purposively varied, was small. Future research could use longitudinal and experimental designs to test whether building AI literacy causally improves calibrated trust and uptake, could combine self-report with analysis of actual revisions and tool logs, and could compare learners across proficiency levels, languages, and institutional settings. Studies that examine how teachers and tools jointly shape feedback, and that include underrepresented languages, would further advance the responsible, inclusive agenda this special issue sets out.

Conclusion:

As AI tools take on more of the work of language assessment, the learner's role does not diminish; it changes. The decisive task becomes judging machine feedback and deciding what to do with it, and this study shows that learners perform that task through calibrated trust, that anxiety born of feedback overload can impede it, and that AI literacy is the capacity that makes it manageable. Trust drove feedback uptake, anxiety dampened it, and AI literacy regulated both, while learners' own accounts revealed a thoughtful, verifying engagement rather than passive acceptance. The implication for CALL-based assessment is clear: the path beyond automation runs through learners who understand the tools well enough to collaborate with them. The Calibrated Trust Framework we propose captures this: the value of an automated assessment is completed in the learner's calibrated decision about what to act on, and AI literacy is what makes that decision sound. Investing in AI literacy, designing for affective load and learner control, and preserving the learner as the decisive agent are the conditions under which AI-mediated assessment becomes pedagogically meaningful, ethically defensible, and genuinely collaborative.

Read together, the four strands of evidence converge on a consistent account. The structural model showed that AI literacy underpins trust and dampens anxiety, that trust is the strongest driver of engagement and uptake, and that anxiety constrains both. The profile analysis showed that these dispositions do not act in isolation but combine into recognisable learner types - Confident Collaborators, Anxious Verifiers, and Detached Low-Literacy Users - each implying a different pedagogical need. The longitudinal analyses showed that AI anxiety is not fixed but declines over a period of AI-mediated collaboration, and that learners with higher AI literacy and trust decline fastest, indicating that literacy shapes not only momentary engagement but the affective trajectory of learning to work with AI. The qualitative accounts explained how this unfolds, revealing verification practices through which learners calibrate rather than simply accept or reject machine feedback. No single strand would support the framework on its own; their convergence is what warrants it.

Recommendations:

Based on these findings, several recommendations follow for practice and policy in EFL writing instruction, particularly in Arab and Gulf higher education. First, institutions and teachers should treat AI literacy as an explicit learning objective rather than an assumed competence, embedding instruction in how to interpret, question, and verify AI-generated feedback into writing curricula. Second, because anxiety from feedback overload was shown to impede uptake, AI feedback tools and tasks should be designed to manage affective load - for example, by staging feedback, foregrounding the most important issues, and preserving learner control over what to act on. Third, the three learner profiles suggest differentiated support: Anxious Verifiers may benefit from reassurance and verification strategies, while Detached Low-Literacy Users need foundational AI literacy and motivational support. Fourth, in contexts where teacher feedback is limited, teacher education should prepare instructors to orchestrate human and AI feedback together rather than cede feedback to tools. Finally, policymakers and tool developers should prioritise transparency and learner agency in AI feedback systems, so that learners remain the decisive agents in assessment. Future research should test whether AI-literacy interventions causally improve calibrated trust and uptake across diverse institutions and languages.

References:

- Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Computer-Assisted Language Learning*, 38(7), 735-331. <https://doi.org/10.1177/07356331241307297>
- Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21, 23. <https://doi.org/10.1186/s41239-024-00455-4>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Chapelle, C. A. (2006). L2 vocabulary acquisition theory: The role of inference, dependability and meaning-focused input. In [Editor(s) needed] (Eds.), *Computer-assisted language learning: [Full book title and page range to be confirmed]* (pp. 33-53). Springer.
- Choung, H., David, P., & Ross, A. (2023). Trust in AI and its influence on the acceptance of AI technologies. *International Journal of Human-Computer Interaction*, 39(9), 1727-1739. <https://doi.org/10.1080/10447318.2022.2050543>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, 32(2), 335-349. <https://doi.org/10.1017/S0272263109990544>
- Elmotri, B., Harizi, R., & Boujlida, A. (2025). The impact of AI-generated feedback explicitness (generic vs. specific) on EFL students' use of automated written corrective feedback. *Arab World English Journal*, 16(1), 384-402. <https://doi.org/10.24093/awej/vol16no1.24>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE Publications.

- Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, 30, 31-44. <https://doi.org/10.1016/j.jslw.2015.08.002>
- Hauck, M. (2025). Learner agency and AI literacy in computer-assisted language learning. ReCALL. Advance online publication.
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125-132. <https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>
- Khasawneh, M. A. S. (2025). Predictive power of AI-generated corrective feedback on EFL learners' self-regulation and language proficiency: Empirical evidence from Saudi Arabian context. *Computer-Assisted Language Learning Electronic Journal*, 26(7), 211-229.
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 100450. <https://doi.org/10.1016/j.asw.2020.100450>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In R. Bernhaupt, F. Mueller, D. Verweij, & J. Andres (Eds.), *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376727>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, 53(4), 914-931. <https://doi.org/10.1111/bjet.13232>
- Pack, A., & Maloney, J. (2024). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *TESOL Quarterly*, 58(2), 1136-1170. <https://doi.org/10.1002/tesq.3253>
- Ranalli, J. (2021). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, 34(7), 936-963. <https://doi.org/10.1080/09588221.2019.1644667>

- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539.
- Stockwell, G. (2024). Technology and motivation in English language teaching and learning. In [Editor(s) needed] (Eds.), *The Routledge handbook of language learning and technology* (pp. [page range needed]). Routledge.
- Voss, E. (2024). Artificial intelligence and language assessment. In *The concise encyclopedia of applied linguistics*. Wiley.
- Wang, Y. Y., & Wang, Y. S. (2022). Development and validation of an artificial intelligence anxiety scale: An initial application in predicting motivated learning behavior. *Interactive Learning Environments*, 30(4), 619-634. <https://doi.org/10.1080/10494820.2019.1674887>
- Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, 10(2), 157-180. <https://doi.org/10.1191/1362168806lr190oa>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90-112. <https://doi.org/10.1111/bjet.13370>
- Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90-102. <https://doi.org/10.1016/j.asw.2018.02.004>