# AI Literacy, Speaking Self-Efficacy, Digital Engagement, and Speaking Anxiety as Predictors of EFL Speaking Performance

**Sutrah[1*], Muhammad Agung[2]**
[1*,2]English Education Study Program at Graduate Program, Faculty of Language and Literature, Universitas Negeri Makassar, Indonesia

**Article Info**
*Article history:*
Received Jun 21, 2026
Accepted Jul 28, 2026
Published Online Aug 28, 2026

***Keywords:***
AI literacy
Speaking self-efficacy
Digital engagement
Speaking anxiety
EFL speaking performance

**ABSTRACT**
Oral proficiency is the area in which Indonesian English majors most often fall short, and the arrival of conversational artificial intelligence has widened access to speaking practice without settling the question of who actually benefits from it. Intervention studies report average gains yet say little about the learner attributes behind them, so departments now weighing AI literacy modules are deciding without evidence on the learners themselves. This study examined how far AI literacy, speaking self-efficacy, digital engagement, and speaking anxiety predict perceived EFL speaking performance, both jointly and individually. A quantitative, non-experimental, cross-sectional correlational survey was conducted with 178 undergraduates in an English Language Education programme at a university in South Sulawesi, drawn by proportionate stratified random sampling with year of study as the stratifying variable. Five Likert scales totalling 63 items measured the four predictors and the outcome, all of them meeting the conventional item and reliability criteria. The questionnaire was administered online over four weeks in mid-semester, and the data were analysed in IBM SPSS Statistics 27 through descriptive statistics, Pearson correlations, classical assumption testing, and simultaneous multiple regression. The four predictors together accounted for 54.6 per cent of the variance in perceived speaking performance, $F(4, 173) = 51.940$, $p < .001$. Speaking self-efficacy made the largest unique contribution ($\beta = .329$, $sr^2 = .068$), followed by speaking anxiety ($\beta = -.252$, $sr^2 = .046$) and digital engagement ($\beta = .217$, $sr^2 = .031$). AI literacy remained significant while contributing least ($\beta = .175$, $sr^2 = .018$), despite a zero-order correlation of .542 with the outcome. Because the design is correlational, that contrast is read as overlap rather than as a causal pathway: the share of variance uniquely attributable to AI literacy shrinks once confidence, practice, and apprehension are taken into account. Departments planning AI literacy modules are therefore better advised to situate them within a wider affective and behavioural strategy than to adopt them as a stand-alone remedy for weak oral performance.


***Corresponding Author:***
Sutrah,
English Education Study Program at Graduate Program,
Faculty of Language and Literature,
Universitas Negeri Makassar, Makassar, Indonesia,
Jalan Andi Pangeran Pettarani, Kelurahan Tidung, Kecamatan Rappocini, Kota Makassar, Provinsi Sulawesi Selatan 90222, Indonesia
Email: *sutrahnorman@gmail.com*

***How to cite:*** Sutrah, S., & Agung, M. (2026). AI Literacy, Speaking Self-Efficacy, Digital Engagement, and Speaking Anxiety as Predictors of EFL Speaking Performance. *Jurnal Riset Dan Inovasi Pembelajaran*, *6*(2), 1129–1152. https://doi.org/10.51574/jrip.v6i2.5740

# *AI Literacy, Speaking Self-Efficacy, Digital Engagement, and Speaking Anxiety as Predictors of EFL Speaking Performance*

## 1. Introduction

Speaking occupies an uncomfortable position in English language education. It is the skill by which employers, examiners, and interlocutors form their first judgement of a graduate, and it is also the skill that learners in foreign language settings develop last and least. Students in English departments across Southeast Asia often handle grammar and reading tasks competently while faltering during a two-minute unscripted turn. The imbalance has structural causes rather than individual ones. Where English is studied as a foreign language, timetabled hours are finite, classes are large, and chances to talk with a proficient interlocutor outside the classroom are rare, so oral practice becomes the part of the curriculum that gets squeezed (Crompton et al., 2024).

Over the past four years that arithmetic has shifted. Generative language models, automatic speech recognition applications, and conversational agents offer learners something classroom scheduling could never supply: a partner available at any hour, indifferent to error, and willing to repeat a task without visible impatience. Bibliometric mapping of intelligent agent research in language education between 2005 and 2025 records a sharp rise in output after 2022, with chatbots and pedagogical agents dominating the recent literature (Huang & Liu, 2026). Controlled comparisons have followed. Fathi et al. (2024) reported that EFL learners practising with a chatbot outperformed a face-to-face group on fluency, lexis, grammatical range, and pronunciation. Yet availability is not the same thing as benefit, and what learners do with these tools varies enormously among students sitting in the same room.

That variation has moved AI literacy from a policy slogan to a measurable learner attribute. Long and Magerko (2020) framed it as the competencies that allow non-specialists to evaluate AI systems, communicate with them, and treat them as collaborators. Later work sharpened the construct for educational use. Ng et al. (2021) distilled the literature into knowing and understanding AI, using and applying it, evaluating and creating with it, and reasoning about its ethics, dimensions subsequently operationalised in a validated questionnaire covering affective, behavioural, cognitive, and ethical facets (Ng et al., 2024). Evidence that these competencies carry consequences is accumulating: among Indian undergraduates, higher AI literacy predicted more purposeful tool use and stronger academic performance through it (Singh et al., 2025), while a survey of 733 Pakistani students found literacy only moderate overall and distributed unevenly across disciplines (Butt et al., 2026).

Speaking self-efficacy rests on much older ground. Bandura (1977) argued that a person's judgement about executing a specific action determines whether the action is attempted, how much effort accompanies it, and how long persistence survives difficulty. He later insisted that such judgements are task-bound rather than global, and that scales measuring them must be written accordingly (Bandura, 1997). Applied to oral English, this concerns a learner's belief about handling a particular speaking demand, not a diffuse impression of being good at the language. The construct has proved sensitive to technology-mediated practice: among 712 Chinese students abroad, ChatGPT use was positively associated with foreign language self-efficacy, with enjoyment carrying part of the relationship (Xu et al., 2024).

Engagement supplies a third strand. Fredricks et al. (2004) consolidated a scattered literature by separating behavioural participation, emotional investment, and cognitive effort, a structure later extended to online settings and adapted specifically for language learners (Abbasi et al., 2024). Digital engagement narrows this to what learners actually do in technology-mediated study, including how often they initiate practice, how absorbed they become while doing it, and how deliberately they work through feedback. The predictive record looks favourable without being uniform. Alshammari and Alrashidi (2024) found engagement

significantly affected EFL achievement in online courses, and Kim et al. (2019) showed academic engagement and digital readiness jointly mediating the link between e-learning perceptions and grade point average. Sun and Zhang (2024) complicate that story: in their Chinese sample, self-perceived engagement correlated with behavioural traces, yet only actual task performance predicted learning outcomes.

Anxiety enters as the counterweight. Horwitz et al. (1986) identified communication apprehension, test anxiety, and fear of negative evaluation as components of a situation-specific anxiety belonging to the language classroom rather than to the learner's disposition, and built the Foreign Language Classroom Anxiety Scale to capture it. MacIntyre and Gardner (1994) subsequently demonstrated that this anxiety consumes attentional resources at the input, processing, and output stages, which helps explain why its effect surfaces most sharply the moment a learner has to speak. A replication of Phillips's oral examination study confirmed the negative association between anxiety and rated oral performance (Hewitt & Stephenson, 2012), regression evidence from postgraduate cohorts indicates that classroom anxiety predicts speaking test scores more strongly than scores on any other skill (M. Liu & Xiangming, 2019), and recent mediation work locates part of the effect in learners' self-appraisal capacity (Ren et al., 2025).

Taken one at a time, these four constructs describe a tidy set of parallel influences. Taken together, they stop behaving. AI literacy turns out not to be affectively neutral. Zhang et al. (2025) surveyed 697 undergraduates using generative AI for English study and found that trust and AI learning self-efficacy mediated the path from tool features to learning engagement, fully in the case of responsiveness and partially in the case of personalisation. Self-efficacy, for its part, operates on anxiety rather than beside it: teacher support reduced AI learning anxiety mainly by way of students' confidence in their own AI use, through a chain running from technophilia to self-efficacy to lowered anxiety (Chai & Sha, 2025). Anxiety and self-efficacy also drift together over time, their negative correlation tightening across a semester instead of holding at a fixed value (Shirvan et al., 2018).

One practical consequence is a measurement problem. Suppose a learner who knows how to prompt an AI tutor also practises more often, feels more capable, and feels less afraid. Any bivariate estimate of the association between AI literacy and speaking performance then carries the other three variables inside it, and the resulting coefficient tells a teacher nothing actionable. Reviews of AI in language education have started to say so. Synthesising twenty-six studies, Ng Kok Wah (2025) concluded that AI systems support engagement and emotional regulation but can also generate technostress and dependency depending on how learners are positioned. A review confined to foreign language anxiety reported chatbots lowering speaking and writing anxiety for most learners while raising it for those who over-rely on them (Pertiwi et al., 2025). Direction of effect appears to hinge on learner attributes, which is exactly what a single-predictor design cannot recover.

Recent syntheses establish that AI tools work in aggregate and leave open why they work so unevenly. Lyu et al. (2025) meta-analysed thirty-one comparative studies and reported a medium effect of chatbots on second language learning ($g = 0.608$), moderated by mobility of access, modality, and whether generative technology was involved. Li et al. (2025) pooled forty-one studies published from 2023 onward, covering 3,515 participants, and found a moderate to large effect on second language acquisition ($ES = 0.576$), with Indonesian as a first language among the strongest moderators. A narrative meta-synthesis spanning eight Asian EFL contexts traced the mechanism to reduced speaking anxiety and non-judgemental feedback while cautioning that evidence for transfer to real-world communication remains thin (Waluyo & Pratiwi, 2025).

Two features of this body of work matter for the present argument. The first is that almost all of it is interventionist. Such designs compare exposure to a tool against conventional

instruction, which answers whether the tool helps on average but says nothing about which learner attributes explain who benefits from it. The second is that null and mixed results cluster in precisely the setting addressed here. Yunita et al. (2025) delivered six AI-assisted speaking sessions to Indonesian undergraduates and found no statistically greater gain in proficiency or willingness to communicate than conventional teaching produced, even though students described reduced anxiety and higher confidence. Qualitative work in Indonesian universities reports the same split between affective gains and stubborn linguistic difficulties (Panggua et al., 2025), and a systematic review of ChatGPT in Indonesian ELT reached a matching verdict about the evidence base itself: qualitative designs predominate, higher education is the usual setting, and the field needs quantitative work capable of modelling relationships rather than cataloguing perceptions (Syairofi et al., 2025).

Stated plainly, the state of the art looks like this. AI literacy is now a validated construct with instruments suited to student populations (Ng et al., 2024). Speaking self-efficacy, digital engagement, and speaking anxiety each carry established measurement traditions and documented links to oral outcomes. Reviews published in the past three years converge on the claim that these variables interact rather than simply accumulate (Huang & Liu, 2026; Ng Kok Wah, 2025).

Three limitations run through that record, and together they mark out the gap this study addresses. The first is structural. The four attributes have grown up in separate literatures and, where they do appear together, they appear in pairs, so no published estimate places all four in one predictive model of speaking performance. The second concerns design. Work on AI and speaking is overwhelmingly interventionist, which establishes whether a tool helps on average but leaves untouched the learner attributes that decide who benefits from it. The third is contextual: the quantitative evidence that does exist comes largely from Chinese, Turkish, and South Asian samples, whereas Indonesian higher education has been examined mainly through perception surveys and qualitative designs (Syairofi et al., 2025). What nobody has yet done is ask what each of the four predictors contributes to speaking performance once the other three are held constant. Without that estimate it cannot be said whether investing in AI literacy training buys anything beyond what confidence building and anxiety reduction already deliver.

For Indonesian English departments the question is not merely academic. Institutions are currently deciding whether to add AI literacy modules to curricula that already struggle to give students enough time to talk, and those decisions rest on intervention studies conducted elsewhere with different learner profiles. If AI literacy predicts speaking performance chiefly because AI-literate students practise more and worry less, the sensible investment lies in engagement and affect. If it predicts performance on its own terms, the case for dedicated instruction becomes considerably stronger.

What the present study does differently is the basis of its claim to novelty. It models AI literacy, speaking self-efficacy, digital engagement, and speaking anxiety simultaneously as predictors of perceived EFL speaking performance, an arrangement absent from the studies reviewed above. It reports the unique contribution of every predictor through squared semipartial correlations, so that the shared and the distinctive portions of each association can be told apart instead of being conflated in a single bivariate coefficient. And it does so in an Indonesian undergraduate setting where the quantitative evidence base is thinnest, using a probability sample rather than the convenience samples that dominate the regional literature.

The objective of this study, accordingly, is to determine how far AI literacy, speaking self-efficacy, digital engagement, and speaking anxiety jointly predict perceived EFL speaking performance among undergraduate English majors, and to estimate what each of them contributes once the remaining three are statistically controlled. Five research questions carry that objective. The first asks how much variance in perceived speaking performance the four predictors explain together. The second asks whether AI literacy retains a significant association

with the outcome after speaking self-efficacy, digital engagement, and speaking anxiety are controlled. The third asks the same of speaking self-efficacy, the fourth of digital engagement, and the fifth of speaking anxiety, in each case with the other three predictors held constant.

## 2. Method

All five research questions ask about the magnitude and the uniqueness of associations among learner attributes that occur naturally rather than being manipulated, so the methodological choices below follow from that. This section sets out the design and the reasoning behind it, the population and sampling procedure, the instruments together with the evidence supporting their quality, the sequence of data collection, and the analytic strategy used to answer each question.

### Type of Research/Design

The study adopted a quantitative, non-experimental, cross-sectional survey design in its correlational and predictive variant. Creswell and Creswell (2017) describe survey research as the appropriate choice when the purpose is to describe trends in a population and to test relationships among variables through a sample studied at one point in time, which corresponds to what the research questions require. No treatment was administered and no group was assigned, because the constructs of interest are attributes that students bring with them into a speaking course rather than conditions an investigator can impose.

Prediction here rests on multiple linear regression. Hair et al. (2019) note that the technique estimates the contribution each independent variable makes to a single metric dependent variable while partialling out variance shared with the remaining predictors, and this property maps directly onto the second through fifth research questions, each of which specifies a control set. The simultaneous entry method was selected in preference to stepwise procedures, since the four predictors are theoretically motivated rather than empirically screened, and stepwise selection capitalises on sample-specific variance in ways that replicate poorly (Tabachnick & Fidell, 2007).

One limitation follows from the design and is acknowledged at the outset. A cross-sectional correlational study supports statements about prediction in the statistical sense, meaning the reduction of uncertainty about an outcome given knowledge of a predictor, but it licenses no inference about causal direction. The reciprocal relationship reported between self-efficacy and anxiety over a semester (Shirvan et al., 2018) makes this caution more than a formality.

### Subjects/Population and Sample

Undergraduate students enrolled in the English Language Education programme of a university in South Sulawesi, Indonesia, formed the target population, provided they had completed at least one credit-bearing speaking course. That criterion was imposed so that every respondent had a shared experiential basis for judging their own oral performance. First-semester students were therefore excluded.

Sample size was determined a priori with G*Power 3.1 (Faul et al., 2007). Specifying a fixed-model linear multiple regression with four predictors, a medium effect size of $f^2 = 0.15$ following Cohen's (1988) conventions, $\alpha = .05$, and statistical power of .95 returned a minimum of 129 cases. The figure was inflated deliberately to absorb attrition and careless responding. Proportionate stratified random sampling was applied with year of study as the stratifying variable, so that second-, third-, and fourth-year cohorts appeared in the sample in the same ratio as in the programme roster; within each stratum, respondents were drawn at random from the departmental enrolment list.

Of 210 students invited, 192 submitted the questionnaire, and 178 cases survived screening for incomplete responses, failed attention checks, and invariant answer patterns, giving a usable response rate of 84.8 per cent.

The retained sample was made up of 128 women (71.9 per cent) and 50 men (28.1 per cent), aged between 18 and 23 years (M = 20.52, SD = 1.31), and divided across the second (n = 69), third (n = 57), and fourth (n = 52) years of study in proportions matching the departmental roster from which the strata were drawn. Self-reported frequency of using AI applications for English learning ranged from rarely (n = 13) to very often (n = 24), with most respondents choosing sometimes (n = 71) or often (n = 70). Table 4 gives the full breakdown. Demographic information covering gender, age, year of study, and frequency of AI use was recorded for descriptive purposes and was not entered into the regression model.

**Instrument**

Data were gathered with a single questionnaire assembling five scales of 63 items in total, each item answered on a five-point Likert format anchored at 1 (strongly disagree) and 5 (strongly agree). Self-report was chosen for the outcome as well as the predictors because the study targets learners' functional command of speaking as they experience it across ordinary academic and social situations, which no single rated task samples adequately.

That choice carries qualifications, and both of them bear on how the results should be read. Since the outcome rests on learners' own appraisals rather than on a rated oral task, it is designated perceived EFL speaking performance throughout, and every statement about it should be read as a statement about self-appraised rather than externally assessed ability; self-ratings of speaking are known to correlate with rated performance without being interchangeable with it. And because predictors and outcome share one response format, common method variance is a genuine threat to the estimates. It was addressed procedurally through the anonymity assurance, the separation of scales within the form, and the randomisation of item order, and statistically through the two tests described under data analysis. Neither remedy has the force of a multi-source design, and the point is returned to among the limitations.

AI literacy was measured with 16 items adapted from the AI Literacy Questionnaire of Ng et al. (2024), retaining its affective, behavioural, cognitive, and ethical structure and rewording contexts so that they referred to English learning rather than to general schooling. The 10 speaking self-efficacy items were written following Bandura's (1997) specification that efficacy scales must be domain-specific and phrased as graded judgements of present capability, so every item takes the form of a can-do statement about a defined oral task. Digital engagement was operationalised through 15 items covering the behavioural, emotional, and cognitive components set out by Fredricks et al. (2004), using the wording developed and validated for online EFL learners by Abbasi et al. (2024).

The 12 speaking anxiety items were drawn from the oral-production subset of the Foreign Language Classroom Anxiety Scale (Horwitz et al., 1986), with the original reverse-worded items preserved to counter acquiescence. Perceived speaking performance was assessed with 10 items spanning fluency, grammatical accuracy, lexical range, pronunciation, and interaction management. The distribution of items across dimensions is set out scale by scale in Table 1.

All items were translated into Indonesian and independently back-translated into English by a second bilingual translator, and discrepancies were reconciled before piloting, in line with the procedure Brislin (1970) established for cross-cultural instruments. Three lecturers in English language teaching judged content relevance and clarity. The revised questionnaire was piloted with 32 students from a comparable programme who did not take part in the main survey.

That pilot was sized for item screening, not for latent variable modelling. It served to check item wording, completion time, and preliminary internal consistency, and it could serve no further purpose: with 32 respondents and 63 items the sample falls well below the case-to-parameter ratios a confirmatory factor analysis of five correlated factors would require.

Evidence on measurement quality was therefore assembled in the main sample through corrected item-total correlations and Cronbach's alpha, reported in full in Tables 2 and 3, and both are treated as evidence of item quality and internal consistency rather than as a substitute for a measurement model. Confirmatory factor analysis judged against the fit criteria summarised by Kline (2023), and discriminant validity assessed through the average-variance-extracted rule of Fornell and Larcker (1981), were not carried out on either sample. Both are recommended for subsequent work on a sample large enough to sustain them, and the omission is listed among the limitations of this study. Table 1 presents the instrument blueprint.

**Table 1.** Instrument Blueprint

| Variable | Dimensions Measured | Items | Sample Item | Adapted From |
|---|---|---|---|---|
| AI Literacy | Affective, behavioural, cognitive, ethical | 16 | I can judge whether the English produced by an AI tool is accurate enough to use. | Ng et al. (2024); Long and Magerko (2020) |
| Speaking Self-Efficacy | Task-specific capability beliefs at graded difficulty | 10 | I can explain my opinion in English for two minutes without preparing notes. | Bandura (1997) |
| Digital Engagement | Behavioural, emotional, cognitive | 15 | I start extra English speaking practice on digital platforms without being asked. | Fredricks et al. (2004); Abbasi et al. (2024) |
| Speaking Anxiety | Communication apprehension, fear of negative evaluation, test anxiety | 12 | I feel my heart pounding when I am called on to speak English in class. | Horwitz et al. (1986) |
| Perceived EFL Speaking Performance | Fluency, accuracy, lexical range, pronunciation, interaction management | 10 | I can keep a conversation in English going without long pauses. | Developed for this study from Hewitt and Stephenson (2012) |

**Procedure/Data Collection**

Ethical approval was obtained from the institutional review committee, and written permission to approach students was granted by the head of the study programme. Participation was voluntary throughout. Every respondent read a consent statement explaining the purpose of the study, the estimated completion time, the absence of any link between responses and course grades, and the right to withdraw at any point before submission.

Distribution took place online across four weeks in the middle of an academic semester, a window chosen to avoid both the settling-in period and the examination weeks, when anxiety readings would be atypical. Collection ran through four steps. Sampling frames were first extracted from the departmental enrolment list and randomised within each year-of-study stratum. Class representatives, briefed beforehand by the researchers, then circulated the questionnaire link inside their cohort groups. Two reminders went out at weekly intervals to students who had not yet responded. Submissions were finally exported to a spreadsheet, screened against the exclusion rules set out below, and only then aggregated into composite

scores. Completion took roughly twenty minutes.

Several safeguards recommended for questionnaire administration in second language research were built into the form (Dörnyei & Taguchi, 2009). Item order was randomised within each scale to disrupt response sets, two attention-check items were embedded at unequal intervals, and no personally identifying information was requested beyond the demographic variables listed earlier. Responses failing either attention check, showing identical answers across an entire scale, or completed in under seven minutes were removed before analysis.

**Data Analysis**

Analysis proceeded in IBM SPSS Statistics version 27 and moved through four stages: measurement quality, description, assumption testing, and hypothesis testing. Measurement quality came first. Corrected item-total correlations were computed for all 63 items and judged against a critical r of .1471 at df = 176 and $\alpha = .05$, and Cronbach's alpha was calculated for each scale against a floor of .70 and an upper guideline of .95, beyond which item redundancy becomes a concern. Descriptive statistics for the five composite variables were computed next, with univariate normality inspected through skewness and kurtosis values and through visual examination of histograms and normal probability plots.

One distinction is worth stating before the assumption checks are described, because the two are easily run together. Univariate normality concerns the distribution of the composite scores themselves, and it was examined for description rather than as a prerequisite of the analysis. Normality of the regression residual concerns the distribution of prediction errors, and it is the assumption that ordinary least squares actually requires. The two were tested separately and are reported separately below.

Assumption testing followed, with a decision rule fixed in advance for each check. Normality of the unstandardised residual was examined with the one-sample Kolmogorov-Smirnov test, the assumption being treated as met where p exceeded .05. Linearity underlying ordinary least squares regression was checked with partial regression plots, and accepted where the plotted relationship showed no systematic curvature. Homoscedasticity was assessed in two ways: by plotting studentised residuals against standardised predicted values, where a random band without funnelling indicates constant error variance, and through the Glejser test, in which no predictor of the absolute residual should reach $p < .05$. Independence of errors was evaluated with the Durbin-Watson statistic against the conventional interval of 1.5 to 2.5. Multicollinearity was judged through tolerance and variance inflation factors, applying the thresholds recommended by Hair et al. (2019) of tolerance above .10 and VIF below 10, with the stricter VIF ceiling of 3 retained as a secondary check appropriate to survey data.

Individual cases were screened last. Mahalanobis and Cook's distances were used to identify multivariate outliers and influential cases (Tabachnick & Fidell, 2007), the former judged against a critical $\chi^2(4)$ of 18.47 at $p < .001$ and the latter against a ceiling of 1.0. Pearson product-moment correlations among all five variables were then examined as a preliminary description of the relational pattern.

Answering the research questions came next. A simultaneous multiple regression entering AI literacy, speaking self-efficacy, digital engagement, and speaking anxiety in one block answered the first question, with the multiple correlation coefficient, $R^2$, adjusted $R^2$, and the omnibus F test indicating how much variance in perceived speaking performance the four predictors jointly explain. The remaining four questions were answered from the same model, since a coefficient in simultaneous regression already represents the association of one predictor with the outcome while the other three are statistically controlled.

For each predictor the unstandardised and standardised coefficients, the t statistic, the exact p value, and the 95 per cent confidence interval are reported, together with the squared semipartial correlation as the proportion of outcome variance uniquely attributable to that predictor and Cohen's (1988) $f^2$ as the accompanying effect size. Confidence intervals were

bootstrapped over 5,000 resamples to reduce dependence on distributional assumptions, and the alpha level was set at .05 throughout. Decisions are expressed in one form throughout the results: where p fell below .05 the null hypothesis was rejected, and where it did not, the null hypothesis was not rejected, a wording that carries no implication that the null hypothesis is true.

A final step examined common method variance, which is a live concern whenever predictors and outcome come from one self-report instrument. Harman's single-factor test was applied first. Its logic is diagnostic rather than corrective. A first unrotated factor accounting for less than half the total variance indicates that no single source dominates the covariance structure, but the test is insensitive to moderate contamination and removes none of it, so a clean result is evidence of nothing worse than the absence of a dominant method factor.

For that reason a full collinearity assessment was added as a second and more sensitive check, in which variance inflation factors above 3.3 in a model regressing every construct on a random dependent variable would signal contamination (Kock, 2015). Both procedural and statistical remedies described by Podsakoff et al. (2003), including the anonymity assurance and the scale separation noted above, were incorporated by design rather than applied after the fact. Neither procedure substitutes for a multi-source design, and the findings below are read with that reservation standing.

## 3. Research Findings

Results are reported in the sequence in which the analysis was run. Measurement quality is established first, because nothing that follows means anything if the scales do not hold. The composition of the sample, the descriptive picture, and the correlation matrix come next, then the assumption checks that licence ordinary least squares estimation, and last the regression model from which all five research questions are answered.

### Item Quality and Internal Consistency

Item quality was examined by correlating every item with the total score of the scale it belonged to. At df = 176 and α = .05 the critical value of r is .1471, so any coefficient above that figure clears the criterion. All 63 items did. Obtained coefficients fell between .598 and .757, each significant at $p < .001$, and the corrected item-total correlations, which remove the item's own contribution to the total, ranged from .539 to .711. Because nothing sat near the threshold, no item was dropped and each scale entered the analysis at full length.

What these coefficients demonstrate is item discrimination and the internal coherence of each scale. They are not on their own evidence of construct validity, which would call for the measurement model described in the method section and not estimable on a pilot of 32 cases. The results below should therefore be read as showing that the items within each scale belong together, not as proof that each scale captures the construct it names. Table 2 sets out the results.

**Table 2.** Item Validity Test Results (Corrected by Item-Total Correlation, N = 178)

| Item | r-count | Sig. | Decision | Item | r-count | Sig. | Decision |
|---|---|---|---|---|---|---|---|
| AIL1 | .669 | .000 | Valid | DE7 | .680 | .000 | Valid |
| AIL2 | .719 | .000 | Valid | DE8 | .704 | .000 | Valid |
| AIL3 | .706 | .000 | Valid | DE9 | .727 | .000 | Valid |
| AIL4 | .686 | .000 | Valid | DE10 | .701 | .000 | Valid |
| AIL5 | .656 | .000 | Valid | DE11 | .714 | .000 | Valid |
| AIL6 | .707 | .000 | Valid | DE12 | .662 | .000 | Valid |
| AIL7 | .757 | .000 | Valid | DE13 | .692 | .000 | Valid |
| AIL8 | .736 | .000 | Valid | DE14 | .695 | .000 | Valid |
| AIL9 | .729 | .000 | Valid | DE15 | .673 | .000 | Valid |

| Item | r-count | Sig. | Decision | Item | r-count | Sig. | Decision |
|---|---|---|---|---|---|---|---|
| AIL10 | .682 | .000 | Valid | ANX1 | .728 | .000 | Valid |
| AIL11 | .688 | .000 | Valid | ANX2 | .698 | .000 | Valid |
| AIL12 | .718 | .000 | Valid | ANX3 | .678 | .000 | Valid |
| AIL13 | .598 | .000 | Valid | ANX4 | .663 | .000 | Valid |
| AIL14 | .698 | .000 | Valid | ANX5 | .720 | .000 | Valid |
| AIL15 | .628 | .000 | Valid | ANX6 | .750 | .000 | Valid |
| AIL16 | .735 | .000 | Valid | ANX7 | .702 | .000 | Valid |
| SSE1 | .714 | .000 | Valid | ANX8 | .749 | .000 | Valid |
| SSE2 | .747 | .000 | Valid | ANX9 | .694 | .000 | Valid |
| SSE3 | .741 | .000 | Valid | ANX10 | .650 | .000 | Valid |
| SSE4 | .735 | .000 | Valid | ANX11 | .731 | .000 | Valid |
| SSE5 | .750 | .000 | Valid | ANX12 | .700 | .000 | Valid |
| SSE6 | .738 | .000 | Valid | SPK1 | .757 | .000 | Valid |
| SSE7 | .731 | .000 | Valid | SPK2 | .669 | .000 | Valid |
| SSE8 | .747 | .000 | Valid | SPK3 | .709 | .000 | Valid |
| SSE9 | .713 | .000 | Valid | SPK4 | .720 | .000 | Valid |
| SSE10 | .726 | .000 | Valid | SPK5 | .698 | .000 | Valid |
| DE1 | .646 | .000 | Valid | SPK6 | .669 | .000 | Valid |
| DE2 | .663 | .000 | Valid | SPK7 | .753 | .000 | Valid |
| DE3 | .672 | .000 | Valid | SPK8 | .708 | .000 | Valid |
| DE4 | .681 | .000 | Valid | SPK9 | .702 | .000 | Valid |
| DE5 | .702 | .000 | Valid | SPK10 | .743 | .000 | Valid |
| DE6 | .677 | .000 | Valid | | | | |

Note. r-table at df = 176 and α = .05 is .1471. An item is retained when r-count exceeds r-table and Sig. < .05. Coefficients index item discrimination and internal consistency, not construct validity. AIL = AI literacy; SSE = speaking self-efficacy; DE = digital engagement; ANX = speaking anxiety; SPK = perceived EFL speaking performance.

Internal consistency was equally untroubled. Cronbach's alpha reached .929 for AI literacy, .919 for digital engagement, .908 for speaking anxiety, .905 for speaking self-efficacy, and .893 for perceived speaking performance. All five values sit far above the .70 floor normally applied to research instruments, and the highest of them stops short of the .95 point at which redundancy among items becomes a worry. Table 3 reports the reliability statistics.

**Table 3.** Reliability Statistics

| Variable | N of Items | Cronbach's Alpha | Interpretation |
|---|---|---|---|
| AI Literacy | 16 | 0.929 | Excellent |
| Speaking Self-Efficacy | 10 | 0.905 | Excellent |
| Digital Engagement | 15 | 0.919 | Excellent |
| Speaking Anxiety | 12 | 0.908 | Excellent |
| Perceived EFL Speaking Performance | 10 | 0.893 | Good |

Note. Alpha computed on the recoded item set; reverse-worded anxiety items were scored 6 − x before analysis.

**Respondent Characteristics and Descriptive Statistics**

Table 4 sets out the composition of the retained sample. Women accounted for 71.9 per cent of respondents, a proportion that mirrors the intake of English education programmes in the region rather than any feature of the sampling procedure. Ages ran from 18 to 23 years (M

= 20.52, SD = 1.31). The three year-of-study strata entered the sample in shares close to their share of the departmental roster, as proportionate stratification was intended to secure. On frequency of AI use, roughly four respondents in five placed themselves in the sometimes or often categories and only 7.3 per cent reported rare use, which suggests that exposure to these tools is now close to universal in this population even though its depth varies.

**Table 4.** Respondent Characteristics (N = 178)

| Characteristic | Category | f | % |
|---|---|---|---|
| Gender | Women | 128 | 71.9 |
| | Men | 50 | 28.1 |
| Year of study | Second year | 69 | 38.8 |
| | Third year | 57 | 32.0 |
| | Fourth year | 52 | 29.2 |
| Frequency of AI use for English learning | Rarely | 13 | 7.3 |
| | Sometimes | 71 | 39.9 |
| | Often | 70 | 39.3 |
| | Very often | 24 | 13.5 |

Note. Percentages are of the 178 retained cases and are rounded to one decimal place. Age ranged from 18 to 23 years (M = 20.52, SD = 1.31). Strata proportions follow the departmental roster used for proportionate stratified random sampling.

Table 5 reports the descriptive statistics for the five composite variables, each expressed on the original five-point metric. AI literacy returned the highest mean (M = 3.732, SD = 0.606) and speaking anxiety the lowest of the four predictors (M = 3.259, SD = 0.694). Interpretation of that anxiety figure rests on the response metric itself rather than on any external benchmark. With 3.00 as the midpoint of the five-point scale, a mean of 3.259 places the sample a little above the neutral point, which is read here as a moderate level of speaking apprehension. No comparison with other populations is offered, since the scale was adapted for this study and no equivalent published mean exists for an Indonesian undergraduate sample. Perceived EFL speaking performance averaged 3.365 (SD = 0.641).

Skewness values lay between -0.326 and -0.002 and kurtosis between -0.779 and -0.055, all within the range that Hair et al. (2019) treat as consistent with univariate normality once the standard errors of .182 and .362 are taken into account. These figures describe the distribution of the composite scores; the normality assumption of the regression concerns the residual and is reported separately below.

**Table 5.** Descriptive Statistics of the Study Variables

| Variable | N | Minimum | Maximum | Mean | Std. Deviation | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| AI Literacy | 178 | 2.188 | 5.000 | 3.732 | 0.606 | -.210 | -.418 |
| Speaking Self-Efficacy | 178 | 1.800 | 5.000 | 3.435 | 0.709 | -.261 | -.411 |
| Digital Engagement | 178 | 2.000 | 4.933 | 3.542 | 0.634 | -.047 | -.779 |
| Speaking Anxiety | 178 | 1.583 | 5.000 | 3.259 | 0.694 | -.002 | -.344 |
| Perceived EFL Speaking Performance | 178 | 1.200 | 4.900 | 3.365 | 0.641 | -.326 | -.055 |

Note. Std. error of skewness = .182; std. error of kurtosis = .362. All scores are scale means on the original five-point response format.

The pattern in Table 6 is what the theoretical argument would lead one to expect, and one detail in it matters for the interpretation that follows. Every predictor correlated significantly with perceived speaking performance: speaking self-efficacy most strongly ($r = .624$, $p < .001$), then speaking anxiety in the negative direction ($r = -.538$, $p < .001$), then AI literacy ($r = .542$, $p < .001$) and digital engagement ($r = .507$, $p < .001$). Predictors were also related to one another. AI literacy and digital engagement shared the largest association ($r = .570$, $p < .001$), while speaking self-efficacy and speaking anxiety moved in opposite directions as Shirvan et al. (2018) reported ($r = -.512$, $p < .001$). Taken at face value, AI literacy looks like the third strongest correlate of perceived speaking performance in the set. The regression qualifies that reading. Every coefficient in the matrix describes an association measured at a single point in time, and none of them supports a statement about which variable acts on which.

**Table 6.** Pearson Product-Moment Correlations Among the Study Variables

| Variable | | AIL | SSE | DE | ANX | SPK |
|---|---|---|---|---|---|---|
| AI Literacy | Pearson Correlation | 1 | .476 | .570 | -.344 | .542 |
| | Sig. (2-tailed) | | .000 | .000 | .000 | .000 |
| Speaking Self-Efficacy | Pearson Correlation | .476 | 1 | .377 | -.512 | .624 |
| | Sig. (2-tailed) | .000 | | .000 | .000 | .000 |
| Digital Engagement | Pearson Correlation | .570 | .377 | 1 | -.261 | .507 |
| | Sig. (2-tailed) | .000 | .000 | | .000 | .000 |
| Speaking Anxiety | Pearson Correlation | -.344 | -.512 | -.261 | 1 | -.538 |
| | Sig. (2-tailed) | .000 | .000 | .000 | | .000 |
| Perceived EFL Speaking Performance | Pearson Correlation | .542 | .624 | .507 | -.538 | 1 |
| | Sig. (2-tailed) | .000 | .000 | .000 | .000 | |

Note. N = 178 for every coefficient. AIL = AI literacy; SSE = speaking self-efficacy; DE = digital engagement; ANX = speaking anxiety; SPK = EFL speaking performance.

## Classical Assumption Testing

Normality of the error term was tested with the one-sample Kolmogorov-Smirnov procedure applied to the unstandardised residual. That statistic came to .051, with an asymptotic two-tailed significance of .731 (Table 7). Since the value exceeds .05, the null hypothesis of a normally distributed residual was not rejected and the assumption is treated as met. Histogram and normal probability plot inspection agreed, showing points clustered along the diagonal without systematic departure.

**Table 7. One-Sample Kolmogorov-Smirnov Test on the Unstandardised Residual**

| | | Unstandardized Residual |
|---|---|---|
| N | | 178 |
| Normal Parameters | Mean | .0000000 |
| | Std. Deviation | .43219334 |
| Most Extreme Differences | Absolute | .051 |
| | Positive | .034 |
| | Negative | -.051 |
| Test Statistic | | .051 |
| Asymp. Sig. (2-tailed) | | .731 |

Note. Test distribution is Normal, calculated from data. Lilliefors significance correction applied. Sig. > .05 indicates a normally distributed residual.

Multicollinearity diagnostics appear in Table 8, and they are judged against the criteria fixed in the method section. Tolerance ranged from .589 for AI literacy to .724 for speaking anxiety, and variance inflation factors from 1.381 to 1.698. Both sets of figures stay comfortably inside the limits of tolerance above .10 and VIF below 10, and every one of them also clears the stricter secondary ceiling of VIF below 3 adopted for survey data. The predictors are correlated, as the theory anticipates, but not to a degree that destabilises the coefficient estimates.

**Table 8.** Multicollinearity Diagnostics

| Model | Tolerance | VIF | Decision |
|---|---|---|---|
| AI Literacy | .589 | 1.698 | No multicollinearity |
| Speaking Self-Efficacy | .626 | 1.597 | No multicollinearity |
| Digital Engagement | .660 | 1.516 | No multicollinearity |
| Speaking Anxiety | .724 | 1.381 | No multicollinearity |

Note. Dependent variable: perceived EFL speaking performance. Criteria fixed in the method section: tolerance > .10 and VIF < 10, with VIF < 3 as a secondary check.

Heteroscedasticity was assessed with the Glejser test, in which the absolute value of the residual is regressed on the four predictors. None of the four coefficients approached significance: AI literacy p = .583, speaking self-efficacy p = .511, digital engagement p = .498, and speaking anxiety p = .990 (Table 9). The omnibus test on this auxiliary regression was likewise non-significant, $F(4, 173) = 0.374$, $p = .827$, with $R^2 = .009$. The scatterplot of studentised residuals against standardised predicted values was inspected alongside the test and agrees with it: the points form a horizontal band roughly between -3 and +3, without funnelling and without any systematic pattern across the range of fitted values. The plot is provided in the supplementary files. Residual variance can therefore be treated as constant across the range of fitted values.

**Table 9.** Glejser Test for Heteroscedasticity

| Model | B | Std. Error | t | Sig. | Decision |
|---|---|---|---|---|---|
| (Constant) | .447 | .220 | 2.036 | .043 | |
| AI Literacy | -.023 | .041 | -.550 | .583 | Homoscedastic |
| Speaking Self-Efficacy | .023 | .034 | .659 | .511 | Homoscedastic |
| Digital Engagement | -.025 | .037 | -.679 | .498 | Homoscedastic |
| Speaking | -.000 | .032 | -.012 | .990 | Homoscedastic |

| Model | B | Std. Error | t | Sig. | Decision |
|---|---|---|---|---|---|
| Anxiety | | | | | |

Note. Dependent variable: absolute value of the unstandardised residual. Model F(4, 173) = 0.374, p = .827, $R^2$ = .009. Sig. > .05 for every predictor indicates constant error variance.

Three further checks completed the assumption set, each judged against the rule fixed beforehand. The Durbin-Watson statistic of 1.797 falls inside the interval of 1.5 to 2.5 that is read as evidence of independent errors. Partial regression plots showed a linear association between each predictor and perceived speaking performance, with no curvature and no sign of a threshold at either end of the predictor range. On the diagnostics for unusual cases, no observation exceeded the critical Mahalanobis value of $\chi^2(4) = 18.47$ at $p < .001$, and no Cook's distance came near the ceiling of 1.0, so no case was influential enough to warrant removal and all 178 observations were retained.

**Multiple Linear Regression, Simultaneous F Test, and Coefficient of Determination**

The first research question asked how much of the variance in perceived speaking performance the four predictors explain together. Table 10 gives the answer. The multiple correlation coefficient was R = .739, and $R^2$ = .546 means that 54.6 per cent of the variability in perceived speaking performance is accounted for by AI literacy, speaking self-efficacy, digital engagement, and speaking anxiety acting jointly. Adjusted $R^2$ of .535 shrinks that figure only slightly, which suggests the model is not being carried by sample-specific noise. Standard error of the estimate came to .437 on a five-point scale.

**Table 10.** Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|---|---|---|---|---|---|
| 1 | .739 | .546 | .535 | .437 | 1.797 |

Note. Predictors: (Constant), AI literacy, speaking self-efficacy, digital engagement, speaking anxiety. Dependent variable: perceived EFL speaking performance.

Whether the model as a whole improves on no model at all was settled by the analysis of variance in Table 11. Regression accounted for 39.705 of the 72.767 total sum of squares, leaving 33.062 as residual, and the resulting ratio of mean squares gave F (4, 173) = 51.940 with $p < .001$. The null hypothesis of no joint prediction was therefore rejected: the four predictors, taken as a block, predict perceived speaking performance at a level that chance does not plausibly explain.

**Table 11.** ANOVA for the Simultaneous Regression Model

| Model | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Regression | 39.705 | 4 | 9.926 | 51.940 | .000 |
| Residual | 33.062 | 173 | .191 | | |
| Total | 72.767 | 177 | | | |

Note. Dependent variable: perceived EFL speaking performance. Sig. reported by SPSS as .000 corresponds to $p < .001$.

Research questions two to five concern the individual predictors with the remaining three held constant, and Table 12 answers all four at once. Every coefficient reached significance, so in each case the null hypothesis was rejected. Speaking self-efficacy was the strongest, B = .298, β = .329, t = 5.088, $p < .001$, with a 95 per cent confidence interval of [.182, .413]. Speaking anxiety followed in the negative direction, B = -.233, β = -.252, t = -4.183, $p < .001$, CI [-.342, -.123]. Digital engagement contributed B = .220, β = .217, t = 3.440, p = .001, CI [.094, .346]. AI literacy remained significant but was the weakest of the four, B = .185, β =

.175, t = 2.617, p = .010, CI [.045, .324]. Bootstrapped intervals over 5,000 resamples reproduced these conclusions, excluding zero for each predictor.

**Table 12.** Regression Coefficients and t Tests

| Model | B | Std. Error | Beta | t | Sig. | 95% CI [LL, UL] | Decision |
|---|---|---|---|---|---|---|---|
| (Constant) | 1.632 | .376 | | 4.339 | .000 | [0.890, 2.374] | |
| AI Literacy | .185 | .071 | .175 | 2.617 | .010 | [0.045, 0.324] | Significant |
| Speaking Self-Efficacy | .298 | .059 | .329 | 5.088 | .000 | [0.182, 0.413] | Significant |
| Digital Engagement | .220 | .064 | .217 | 3.440 | .001 | [0.094, 0.346] | Significant |
| Speaking Anxiety | -.233 | .056 | -.252 | -4.183 | .000 | [-0.342, -0.123] | Significant |

Note. Dependent variable: perceived EFL speaking performance. B = unstandardised coefficient; Beta = standardised coefficient. Bootstrapped 95 per cent intervals over 5,000 resamples: AI literacy [0.036, 0.331]; speaking self-efficacy [0.201, 0.389]; digital engagement [0.096, 0.344]; speaking anxiety [-0.346, -0.125].

Reading the squared semipartial correlations alongside the zero-order correlations shows where the model does its real work. Speaking self-efficacy uniquely explained 6.8 per cent of the variance in perceived speaking performance, speaking anxiety 4.6 per cent, and digital engagement 3.1 per cent. AI literacy uniquely explained 1.8 per cent, with an accompanying effect size of $f^2 = .040$, which by Cohen's (1988) benchmarks is small. Set against its zero-order correlation of .542, that unique share is modest, and the drop from a bivariate r of .542 to a partial correlation of .195 marks the difference between what AI literacy looks like on its own and what remains of it once confidence, practice, and apprehension are in the equation. The comparisons drawn in the discussion rest on these squared semipartial values, which are proportions of variance and can be compared directly, rather than on the standardised coefficients, which cannot.

**Table 13.** Unique Contribution of Each Predictor

| Predictor | Zero-order r | Partial r | Part r | $sr^2$ | $f^2$ | Effect Size |
|---|---|---|---|---|---|---|
| AI Literacy | .542 | .195 | .134 | .018 | .040 | Small |
| Speaking Self-Efficacy | .624 | .361 | .261 | .068 | .150 | Medium |
| Digital Engagement | .507 | .253 | .176 | .031 | .068 | Small to medium |
| Speaking Anxiety | -.538 | -.303 | -.214 | .046 | .101 | Small to medium |

Note. $sr^2$ is the squared part (semipartial) correlation, read as the proportion of variance in perceived speaking performance uniquely attributable to the predictor. Effect sizes follow Cohen's (1988) benchmarks of .02 (small), .15 (medium), and .35 (large) for $f^2$.

In unstandardised form the fitted model is: Speaking Performance = 1.632 + 0.185 (AI Literacy) + 0.298 (Speaking Self-Efficacy) + 0.220 (Digital Engagement) - 0.233 (Speaking Anxiety). A one-point rise on the speaking self-efficacy scale is thus associated with roughly a

third of a point on the speaking performance scale, other things being equal, whereas a one-point rise in anxiety costs slightly less than a quarter of a point.

**Common Method Variance**

Because predictors and outcome came from the same self-report instrument, two checks were run before the results were accepted. Unrotated exploratory factor analysis of all 63 items returned 11 factors with eigenvalues above 1, the first of which accounted for 28.94 per cent of the total variance, well under the 50 per cent that Harman's procedure treats as a warning sign. The full collinearity assessment recommended by Kock (2015) was applied as a second and more sensitive test; the highest variance inflation factor obtained was 2.117, below the 3.3 criterion. Taken together the two results give no grounds for attributing the observed relationships to shared method variance. They do not, however, establish that the estimates are free of it, since Harman's test detects only a dominant method factor and the collinearity assessment only a pervasive one; the caution registered in the method section stands.

## 4. Discussion

The model answers the question the introduction posed, and it answers it with a twist. Four learner attributes together explain 54.6 per cent of the variance in perceived speaking performance, a substantial share for a set of psychological and behavioural variables in an applied linguistics survey. That figure settles the first research question. It does not settle the other four, because the predictors do not contribute in proportion to how impressive each looks on its own. AI literacy is the clearest case. Its bivariate correlation with perceived speaking performance was the third highest in the matrix, yet once the other three predictors were entered its unique contribution fell to under two per cent of outcome variance. Everything of interest in this study sits in that gap.

**AI Literacy Predicts Less Uniquely Than It Appears To**

The shrinkage from $r = .542$ to $\beta = .175$ is not a failure of the construct. It is a statement about how much of the association it shares with the other three attributes. AI-literate students in this sample were also the students who practised more on digital platforms ($r = .570$), rated themselves as more capable speakers ($r = .476$), and reported less apprehension about speaking ($r = -.344$). The association between AI literacy and perceived speaking performance therefore overlaps considerably with those attributes, and the unique contribution that survives once they are accounted for is small although still statistically detectable. That is the answer to the second research question: AI literacy retains a significant partial association with the outcome, but a modest one.

Findings from adjacent literatures show a comparable pattern. Studies using mediation frameworks report that the association between AI-related competence and academic outcomes is largely shared with psychological and behavioural variables rather than standing apart from them (Dewi & Alam, 2025; Yi & Siquan, 2025; Yuan et al., 2025). Dewi and Alam (2025) surveyed 452 Indonesian university students and found learning resilience fully mediating the paths from AI literacy and digital self-efficacy to grade point average, which is the closest available parallel to the present sample. Duong (2026) traced the same architecture outside education, where digital entrepreneurial self-efficacy mediated the path from AI literacy to entrepreneurial intention among 1,061 Vietnamese students.

The present result parts company with those studies on one point that is worth stating precisely. Full mediation was not found here, and could not have been, because no mediation model was estimated. What the data show is that AI literacy kept a significant partial association with perceived speaking performance, small though it was, which suggests that the capacity to evaluate and interrogate an AI system is not wholly reducible to practising more or worrying less. Singh et al. (2025) reached a compatible conclusion for Indian undergraduates, where AI literacy predicted purposeful tool use and academic performance through it. A contrasting

picture comes from Pakistan, where literacy was only moderate and unevenly spread across disciplines (Butt et al., 2026), which indicates that the size of this residual association probably depends on how much variation in AI literacy a given population actually contains.

**Speaking Self-Efficacy as the Strongest Single Predictor**

Speaking self-efficacy accounted for more unique variance than any other predictor in the model, and this is the finding with the deepest theoretical backing. Bandura (1997) argued that task-specific capability judgements govern whether an action is attempted at all, how much effort it receives, and how long it survives difficulty. An oral task in a foreign language is precisely the kind of performance where those three things determine the outcome, since a student who abandons a turn after the first hesitation produces measurably less language than one who pushes past it.

Correlational studies in EFL settings have documented the association repeatedly (Hermagustiana et al., 2021; Okyar, 2023; Xu et al., 2024). Okyar (2023) found speaking self-efficacy accounting for 26.7 per cent of the variance in speaking anxiety among 293 Turkish EFL students, a figure that captures how tightly the two constructs are bound. Hermagustiana et al. (2021) obtained a positive correlation between self-efficacy and rated speaking performance among Indonesian sixth-semester students together with a negative one between anxiety and performance, the same double pattern observed here.

What this study adds is the comparison. Placed in the same equation, speaking self-efficacy uniquely accounts for 6.8 per cent of the variance in perceived speaking performance against 1.8 per cent for AI literacy, a ratio of roughly 3.8 to 1. That figure is computed from the squared semipartial correlations, which are proportions of variance and therefore divisible, and not from the standardised coefficients, which are not. The third research question is answered accordingly: speaking self-efficacy retains the largest unique association with the outcome once the other three predictors are controlled. That ordering has a practical reading. Confidence in one's own speaking is not a by-product of technological competence, and treating it as one would misallocate instructional effort.

**Digital Engagement and the Question of What Learners Actually Do**

Digital engagement made a significant unique contribution of 3.1 per cent, which places it third in the model and answers the fourth research question affirmatively. That figure aligns with a body of work linking engagement to language achievement in technology-mediated settings (Alshammari & Alrashidi, 2024; Istiara et al., 2025; Kim et al., 2019). Istiara et al. (2025) applied latent profile analysis to 300 Indonesian English majors and found the most deeply engaged profile scoring highest on both academic self-efficacy and language performance, with the least engaged profile at the bottom on both. Jin (2024) approached the question from the task side, reporting that a social media-integrated vlogging activity raised speaking proficiency while altering learners' affective response, which suggests that what learners engage with matters as much as how much they engage.

There is a caution attached. Sun and Zhang (2024) reported that self-perceived engagement correlated with behavioural traces but that only actual task performance predicted learning outcomes, which raises a legitimate question about what a self-report engagement scale measures. This study used self-report throughout and cannot settle that question. It can note that the engagement coefficient survived the entry of self-efficacy and anxiety, meaning the association is not simply confident students describing themselves as engaged.

Liu and Zhao (2026) offer a useful reframing from the other direction. Analysing 1,012 Chinese EFL learners with hierarchical regression, they found English learning confidence and digital competence predicting participation in AI-mediated informal digital learning. Set beside the present findings, the two studies describe a cluster of mutually associated attributes rather than a sequence: confidence and competence go together with digital practice, and digital practice goes together with stronger self-appraised performance. Which of them precedes the

others is a question neither cross-sectional design can answer.

**Speaking Anxiety as the Counterweight**

Speaking anxiety was the second strongest predictor and the only negative one, uniquely explaining 4.6 per cent of the variance and answering the fifth research question in the affirmative. MacIntyre and Gardner (1994) gave the mechanism: anxiety consumes attentional resources at input, processing, and output, and speaking is where all three stages are compressed into real time. Its negative association with oral performance has been confirmed across contexts and instruments (Hewitt & Stephenson, 2012; Janitra & Anwar, 2022; M. Liu & Xiangming, 2019; Ren et al., 2025). Janitra and Anwar (2022) reported a correlation of -.702 between FLCAS scores and final speaking marks among Indonesian English education students, a stronger figure than the -.538 obtained here, a difference that most likely reflects the outcome measure, since theirs was a lecturer-assigned grade and the present one a self-appraisal.

The AI-related literature complicates the picture in a way this study cannot resolve. Yang et al. (2026) found multimodal AI reducing anxiety and boredom while raising enjoyment among EFL speakers, and Shi and Shakibaei (2025) reported AI-integrated speaking instruction lowering shyness and demotivation. Against that, chatbots have also been found to raise anxiety among learners who over-rely on them (Pertiwi et al., 2025). A cross-sectional design measures the level of anxiety a student brings into the room; it says nothing about which tool produced it. Whether AI exposure is reducing anxiety for most of this sample while amplifying it for a minority is a question for a longitudinal design.

**Implications for Indonesian English Departments**

The introduction framed a decision that Indonesian programmes are currently facing: whether to add AI literacy modules to curricula already short of speaking time. These results argue for a qualified yes with a clear condition attached. AI literacy does retain an independent association with perceived speaking performance, so instruction in it is unlikely to be wasted. But that association is small, and the three attributes it overlaps with are larger. A module teaching prompting technique without also building capability beliefs and reducing apprehension would be addressing the weakest of the four predictors. This is a recommendation read off a predictive pattern, not a demonstration that AI literacy instruction raises oral performance; only an intervention design could establish the latter.

This reading fits the null findings that prompted the study. Yunita et al. (2025) delivered six AI-assisted speaking sessions to Indonesian undergraduates and recorded no greater proficiency gain than conventional teaching produced, even though students described feeling more confident. The same split between affective gains and persistent linguistic difficulty turns up in qualitative work from Indonesian universities (Panggua et al., 2025). If the predictive weight sits with self-efficacy and anxiety rather than with tool competence, a short intervention that changes how students feel without changing how much they speak would plausibly produce that pattern, although the present design can do no more than make the reasoning consistent with the data.

Evidence from studies that did produce measurable gains points the same way. Sok and Shin (2025) recorded improvements in oral ability after sustained interaction with ChatGPT, and Wang et al. (2025) reported gains across speaking dimensions in which willingness to communicate emerged as the decisive factor. Adoption of ChatGPT for speaking practice has been traced to perceived usefulness and ease of use rather than to technical skill (Zou et al., 2025). Across these studies the variable associated with movement in performance is the one governing whether the learner keeps talking, which is what the self-efficacy coefficient in Table 12 represents.

For curriculum design the sequence therefore matters more than the content list. Syairofi et al. (2025) concluded from a systematic review of ChatGPT in Indonesian ELT that the field

has accumulated perceptions and needs models. The model reported here is one such contribution, and, so far as the literature reviewed above records, the first to estimate the unique share of variance each of these four attributes explains within a single Indonesian sample. Its practical message is that AI literacy instruction is better embedded in speaking courses that already work on confidence and apprehension than delivered as a separate technical unit.

**Limitations**

Five constraints bound these conclusions. The design is cross-sectional, so the coefficients describe prediction in the statistical sense and carry no information about causal direction; the reciprocal drift between self-efficacy and anxiety over a semester (Shirvan et al., 2018) makes that caution substantive rather than formulaic. All five variables came from one self-report instrument, and although Harman's test and the full collinearity assessment found no evidence of method contamination, a rated speaking task would have been a stronger outcome measure, which is why the outcome is described throughout as perceived speaking performance. Measurement evidence was confined to item analysis and internal consistency: the pilot of 32 cases could not support a confirmatory factor analysis, so no factor loadings, fit indices, or average-variance-extracted values are available for these scales, and the construct validity claim is correspondingly limited.

The remaining constraints concern how far the findings reach rather than how sound they are inside the sample. The sample was drawn from a single programme in South Sulawesi and was predominantly female, which limits generalisation to other institutions and regions. The model also treats the four predictors as parallel, whereas the literature reviewed above suggests they may be arranged in sequence; a mediation or path model on comparable data would test that arrangement directly.

## 5. Conclusion

This study set out to establish how far AI literacy, speaking self-efficacy, digital engagement, and speaking anxiety predict perceived EFL speaking performance among university students, and what each contributes once the other three are held constant. Taken together the four attributes explain 54.6 per cent of the variance in perceived speaking performance, and the model is significant as a whole. Each predictor also retained a significant partial association, so the answer to all five research questions is affirmative in the narrow sense that every variable earns its place in the equation.

The ordering within that model is where the substance lies. Speaking self-efficacy makes the largest unique contribution, speaking anxiety the second largest in the opposite direction, and digital engagement the third. AI literacy comes fourth by a considerable margin, explaining under two per cent of outcome variance on its own despite a bivariate correlation of .542 with the outcome. That contrast between the simple and the partial estimate is the study's central result. What it shows is that the unique contribution of AI literacy is substantially reduced after speaking self-efficacy, digital engagement, and speaking anxiety are controlled, which is to say that most of what the bivariate coefficient registers is shared with those three attributes. Whether AI literacy operates through them is a different question, one this design neither tested nor was built to test.

Three recommendations follow from what the model predicts rather than from any demonstrated effect. Departments weighing whether to introduce AI literacy modules should proceed, but should embed them in speaking courses instead of delivering them as stand-alone technical units, since the attributes carrying most of the predictive weight are addressed there. Instructional design should give priority to raising students' belief in their capacity to hold an extended turn and to lowering their fear of being evaluated while doing so. And lecturers puzzled by a technologically capable student who still speaks poorly have reason to look first at confidence and apprehension, which this model identifies as the stronger correlates of

perceived performance.

Four directions for further work follow from the design's limits. A longitudinal study would establish whether AI literacy accompanies or precedes speaking self-efficacy over time, and a mediation model on comparable data would test the sequential arrangement the correlations here imply. Replacing the self-report outcome with a rated oral task would remove the shared-method concern that no statistical remedy fully settles. A validation study on a sample large enough for confirmatory factor analysis would supply the factor loadings, fit indices, and discriminant validity evidence that the present pilot could not. Sampling across several Indonesian institutions with differing levels of technological provision would show whether the small unique association of AI literacy observed here grows in populations where the construct varies more widely.

## 6. Conflict of Interest

The authors declare no conflict of interest.

## 7. Author Contributions

S. and M.A. are postgraduate students who collaboratively conceptualized the research idea and developed the theoretical framework. Both authors jointly designed the methodology, developed the research instruments, and coordinated the data collection process. S. performed the statistical analysis using IBM SPSS Statistics 27 and drafted the initial manuscript, while M.A. contributed to data organization, interpretation of the results, and critical revision of the manuscript. Both authors (S. and M.A.) actively participated in the discussion of the findings and approved the final version of the work. All authors confirm that they have read and approved the final version of this manuscript. The percentage contributions for the conceptualization, data collection, analysis, drafting, and revision of this paper are as follows: S.: 50%, and M.A.: 50%.

## 8. Data Availability Statement

The data supporting the findings of this study are available as “supplementary files”.

## REFERENCES

Abbasi, M., Ghamoushi, M., & Mohammadi Zenouzagh, Z. (2024). EFL learners’ engagement in online learning context: Development and validation of potential measurement inventory. *Universal Access in the Information Society*, *23*(3), 1467–1481. https://doi.org/10.1007/s10209-023-00993-0

Alshammari, S. H., & Alrashidi, O. (2024). The Effect of Students’ Engagement on Their Learning Achievement in EFL Online Courses: A Structural Equation Modelling Approach. *International Journal of Online Pedagogy and Course Design*, *14*(1), 1–18. https://doi.org/10.4018/IJOPCD.357875

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Brislin, R. W. (1970). Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Butt, F. S., Malik, A., & Mahmood, K. (2026). Assessing AI literacy of university students in Pakistan: A cross-sectional survey. *Digital Library Perspectives*, *42*(1), 141–158. https://doi.org/10.1108/DLP-06-2025-0094

Chai, T., & Sha, C. (2025). Prevention of AI learning anxiety: The chain mediation role of technophilia and AI self-efficacy under the context of teacher support. *Acta Psychologica*, *261*, 105979. https://doi.org/10.1016/j.actpsy.2025.105979

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, *55*(6), 2503–2529. https://doi.org/10.1111/bjet.13460

Dewi, E. R., & Alam, A. A. (2025). AI Literacy, Digital Self-Efficacy, and Learning Resilience as Predictors of Academic Success Across Urban and Rural Students. *Journal of Educational Science and Technology (EST)*, *11*(2), 60. https://doi.org/10.26858/est.v11i2.75877

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in Second Language Research: Construction, Administration, and Processing* (2nd ed.). Routledge. https://doi.org/10.4324/9780203864739

Duong, C. D. (2026). AI literacy and higher education students' digital entrepreneurial intention: A moderated mediation model of AI self-efficacy and digital entrepreneurial self-efficacy. *Industry and Higher Education*, *40*(2), 242–255. https://doi.org/10.1177/09504222251370089

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, *121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis*. Andover, UK: Cengage Learning. https://books.google.co.id/books?id=0R9ZswEACAAJ

Hermagustiana, I., Astuti, A. D., & Sucahyo, D. (2021). Do I Speak Anxiously? A Correlation of Self-Efficacy, Foreign Language Learning Anxiety and Speaking Performance of Indonesian EFL Learners. *Script Journal: Journal of Linguistics and English Teaching*, *6*(1), 68–80. https://doi.org/10.24903/sj.v6i1.696

Hewitt, E., & Stephenson, J. (2012). Foreign Language Anxiety and Oral Exam Performance: A Replication of Phillips's *MLJ* Study. *The Modern Language Journal*, *96*(2), 170–189. https://doi.org/10.1111/j.1540-4781.2011.01174.x

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

Huang, L., & Liu, M. (2026). The Rise of Digital Tutors: A Bibliometric Exploration of Intelligent Agent-Supported Language Learning From 2005 to 2025. *European Journal of Education*, *61*(1), e70529. https://doi.org/10.1111/ejed.70529

Istiara, F., Ab, J. S., Tama, O. H., Mahrunnisya, D., Ajeng, G. D., Adijaya, N., Saputra, D. W., & Hadi, M. S. (2025). Deep Learning Engagement as a Predictor of Academic Self-Efficacy and Language Performance. *TARBIYA: Journal of Education in Muslim Society*, 115–126. https://doi.org/10.15408/tjems.v12i1.46710

Janitra, M. J. O., & Anwar, K. (2022). Significant Relationship between Anxiety Level and Oral Communication Performances: The Case of English Education Learners. *New Language Dimensions*, *3*(1), 60–70. https://doi.org/10.26740/nld.v3n1.p60-70

Jin, S. (2024). Speaking proficiency and affective effects in EFL: Vlogging as a social media-integrated activity. *British Journal of Educational Technology*, *55*(2), 586–604. https://doi.org/10.1111/bjet.13381

Kim, H. J., Hong, A. J., & Song, H.-D. (2019). The roles of academic engagement and digital readiness in students' achievements in university e-learning environments. *International Journal of Educational Technology in Higher Education*, *16*(1), 21. https://doi.org/10.1186/s41239-019-0152-3

Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling*. Guilford Publications. https://books.google.co.id/books?id=t2CvEAAAQBAJ

Kock, N. (2015). Common Method Bias in PLS-SEM: A Full Collinearity Assessment Approach. *International Journal of e-Collaboration*, *11*(4), 1–10. https://doi.org/10.4018/ijec.2015100101

Li, M., Wang, Y., & Yang, X. (2025). Can Generative AI Chatbots Promote Second Language Acquisition? A Meta-Analysis. *Journal of Computer Assisted Learning*, *41*(4), e70060. https://doi.org/10.1111/jcal.70060

Liu, G. L., & Zhao, X. (2026). The Predictive Effects of Sociobiographical Variables, English Learning Confidence, and Digital Competence on AI-Mediated Informal Digital Learning of English (AI-IDLE). *International Journal of Applied Linguistics*, *36*(2), 1641–1652. https://doi.org/10.1111/ijal.70010

Liu, M., & Xiangming, L. (2019). Changes in and Effects of Anxiety on English Test Performance in Chinese Postgraduate EFL Classrooms. *Education Research International*, *2019*, 1–11. https://doi.org/10.1155/2019/7213925

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3313831.3376727

Lyu, B., Lai, C., & Guo, J. (2025). Effectiveness of Chatbots in Improving Language Learning: A Meta-Analysis of Comparative Studies. *International Journal of Applied Linguistics*, *35*(2), 834–851. https://doi.org/10.1111/ijal.12668

MacIntyre, P. D., & Gardner, R. C. (1994). The Subtle Effects of Language Anxiety on Cognitive Processing in the Second Language. *Language Learning*, *44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology*, *55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

Ng Kok Wah, J. (2025). Artificial Intelligence in Language Learning: A Systematic Review of Personalization and Learner Engagement. *Forum for Linguistic Studies*, *7*(9). https://doi.org/10.30564/fls.v7i9.10336

Okyar, H. (2023). Foreign Language Speaking Anxiety and its Link to Speaking Self-Efficacy, Fear of Negative Evaluation, Self-Perceived Proficiency and Gender. *Science Insights Education Frontiers*, *17*(2), 2715–2731. https://doi.org/10.15354/sief.23.or388

Pongsapan, N. P., Ismail, H., Patandung, Y., Rachel, & Tangirerung, J. R. (2025). AI-Enhanced Academic Speaking Skills: A Qualitative Investigation of Digital Tool Integration in Indonesian EFL University Context. *Forum for Linguistic Studies*, *7*(8).

https://doi.org/10.30564/fls.v7i8.10763

Pertiwi, P. C., Segoh, D., & Rohmadhani, A. (2025). Artificial Intelligence in Academic Environments: Reducing or Increasing Foreign Language Anxiety? *Inspiring: English Education Journal*, *8*(1), 100–131. https://doi.org/10.35905/inspiring.v8i1.13009

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Ren, J., Nik, W. S. B. W., Wang, S., & Wang, H. (2025). Language evaluation anxiety and self-appraisal ability: Their impact on students' oral English performance. *Environment and Social Psychology*, *10*(7). https://doi.org/10.59429/esp.v10i7.3905

Shi, W., & Shakibaei, G. (2025). Insights Into the Effectiveness of Artificial Intelligence-Integrated Speaking Instruction in Enhancing Speaking Skills and Social–Emotional Competence as Well as Reducing Demotivation and Shyness. *European Journal of Education*, *60*(3), e70174. https://doi.org/10.1111/ejed.70174

Shirvan, M. E., Khajavy, G. H., Nazifi, M., & Taherian, T. (2018). Longitudinal examination of adult students' self-efficacy and anxiety in the course of general English and their prediction by ideal self-motivation: Latent growth curve modeling. *New Horizons in Adult Education and Human Resource Development*, *30*(4), 23–41. https://doi.org/10.1002/nha3.20230

Singh, E., Vasishta, P., & Singla, A. (2025). AI-enhanced education: Exploring the impact of AI literacy on generation Z's academic performance in Northern India. *Quality Assurance in Education*, *33*(2), 185–202. https://doi.org/10.1108/QAE-02-2024-0037

Sok, S., & Shin, H. W. (2025). Do Interactions with CHATGPT Influence L2 Learners' Oral Speaking Ability, Summarization Ability, and Perceptions of Generative AI Tasks? *TESOL Quarterly*, *59*(S1). https://doi.org/10.1002/tesq.70001

Sun, P. P., & Zhang, L. J. (2024). Investigating the Effects of Chinese University Students' Online Engagement on Their EFL Learning Outcomes. *The Asia-Pacific Education Researcher*, *33*(4), 747–757. https://doi.org/10.1007/s40299-023-00800-7

Syairofi, A., Emilia, E., & Purnawarman, P. (2025). The use of ChatGPT in the Indonesian ELT context: A systematic review. *Journal of Research on English and Language Learning (J-REaLL)*, *6*(2), 217–234. https://doi.org/10.33474/j-reall.v6i2.23552

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon/Pearson Education.

Waluyo, B., & Pratiwi, D. I. (2025). AI chatbot-assisted English learning and willingness to communicate: A narrative meta-synthesis of evidence from Asian English as a foreign language contexts. *The JALT CALL Journal*, *21*(3), 102884. https://doi.org/10.29140/jaltcall.v21n3.102884

Wang, C., Li, X., & Zou, B. (2025). Revisiting Integrated Model of Technology Acceptance Among the Generative AI -Powered Foreign Language Speaking Practice: Through the Lens of Positive Psychology and Intrinsic Motivation. *European Journal of Education*, *60*(1), e70054. https://doi.org/10.1111/ejed.70054

Xu, S., Chen, P., & Zhang, G. (2024). Exploring the impact of the use of ChatGPT on foreign language self-efficacy among Chinese students studying abroad: The mediating role of foreign language enjoyment. *Heliyon*, *10*(21), e39845. https://doi.org/10.1016/j.heliyon.2024.e39845

Yang, C., Shao, J., Zhang, M., & Xu, H. (2026). Integrating Multimodality in Emotion Moderation: The Effects of Multimodal AI on EFL Speakers' Boredom, Enjoyment, and Anxiety. *International Journal of Applied Linguistics*, ijal.70201. https://doi.org/10.1111/ijal.70201

Yi, Z., & Siquan, X. (2025). The Impact of Pre-Service Language Teachers' Basic Psychological Needs on Behavioural Intentions to Utilise Artificial Intelligence ( AI ) in Teaching: AI Literacy and Self-Efficacy as Mediators. *European Journal of Education*, *60*(3), e70160. https://doi.org/10.1111/ejed.70160

Yuan, N., Yu, Q., & Liu, W. (2025). The impact of digital literacy on learning outcomes among college students: The mediating effect of digital atmosphere, self-efficacy for digital technology and digital learning. *Frontiers in Education*, *10*, 1641687. https://doi.org/10.3389/feduc.2025.1641687

Yunita, R., Febrilyantri, C., Hayasaki, A., & Ni'am, M. I. (2025). Bridging the Affective-Linguistic Gap: A Mixed-Methods Exploration of AI-Assisted Speaking Practice and Willingness to Communicate. *Jurnal Pendidikan Progresif*, *15*(4), 2760–2780. https://doi.org/10.23960/jpp.v15i4.pp2760-2780

Zhang, Q., Pan, X., & Fan, J. (2025). Uncovering the Mediating Effects of AI Trust and Self-Efficacy on Engagement in Human- GENAI Communication for EFL Learning. *European Journal of Education*, *60*(4), e70336. https://doi.org/10.1111/ejed.70336

Zou, B., Wang, C., Yan, Y., Du, X., & Ji, Y. (2025). Exploring English as a Foreign Language Learners' Adoption and Utilisation of ChatGPT for Speaking Practice Through an Extended Technology Acceptance Model. *International Journal of Applied Linguistics*, *35*(2), 689–704. https://doi.org/10.1111/ijal.12658

**Author Biographies**


| | |
|---|---|
|  | **Sutrah,** is a postgraduate student in the English Education Study Program at the Graduate Program, Faculty of Language and Literature, Universitas Negeri Makassar, Indonesia. Her research interests lie in the field of English Language Teaching (ELT), with particular focus on EFL Speaking Skills, Technology-Enhanced Language Learning (TELL), Artificial Intelligence in Education (AIED), and online and Digital Learning Platforms. She can be contacted at: sutrahnorman@gmail.com or ORCID: https://orcid.org/0009-0001-8889-3717 |
|  | **Muhammad Agung,** is a student in the Graduate Program of English Education at Universitas Negeri Makassar, Indonesia. He completed his teacher professional education (PPG Prajabatan) in 2023 through a scholarship from the Ministry of Education and Culture. He has an interest in eLearning creation, exemplified by one of his released learning applications on the Google Play Store, which applying gamification to English vocabulary learning media for Junior High School level named Lima Sahabat (https://play.google.com/store/apps/details?id=com.ajaran.limasahabat). |