Journal of Applied Studies in Language, Volume 10 Issue 1 (June 2026), p. 16-25
p-issn 2598-4101 e-issn 2615-4706 © Politeknik Negeri Bali
http://ojs2.pnb.ac.id/index.php/JASL

The Effect of Gemini AI-enhanced E-portfolios on Students' Speaking Performance: A Quantitative Analysis in ESP Context

I Putu Yoga Laksana[1,2], Ni Made Ratminingsih[3], Made Hery Santosa[4], I Putu Indra Kusuma[5], Thijs Hemmer[6]

Universitas Pendidikan Ganesha, Indonesia[1,3,4,5]
Politeknik Negeri Bali, Indonesia[2]
Windesheim University, Netherland[6]
email: yoga.laksana@pnb.ac.id[2]

Abstract - This study investigates the quantitative impact of Gemini AI-enhanced e-portfolio assessment on English speaking performance among ESP students in a vocational education context. Despite the growing integration of artificial intelligence in language education, empirical evidence examining the effectiveness of multimodal AI systems on speaking skill development remains limited, particularly within ESP contexts where students require targeted professional language competencies. Using a one-group pretest-posttest design, 30 second-semester International Business Management students at Politeknik Negeri Bali participated in a 14-week intervention utilizing Gemini AI-integrated e-portfolio assessment. Speaking performance was measured through comprehensive assessments based on the Complexity, Accuracy, and Fluency (CAF) framework, incorporating formal business presentations and structured role-play negotiation scenarios representative of authentic professional discourse demands. Results from paired samples t-test analysis revealed statistically significant improvements following the intervention ($t(29) = 15.847$, $p < 0.001$, Cohen's $d = 2.89$), with mean speaking scores increasing substantially from 62.45 ($SD = 8.23$) to 78.67 ($SD = 6.45$). Component-level analysis demonstrated differential yet consistently strong effects across all CAF dimensions: fluency ($d = 2.52$), complexity ($d = 2.27$), and accuracy ($d = 1.87$). The very large effect sizes indicate substantial practical significance, exceeding effect sizes reported in previous technology-enhanced speaking interventions by factors of four to eight. These findings contribute critical empirical evidence supporting AI integration in vocational language education and suggest that Gemini AI-enhanced e-portfolio assessment represents a paradigm shift in ESP pedagogy, with important implications for curriculum redesign and institutional technology adoption decisions in professional preparation programs.

Keywords: CAF framework, E-portfolio assessment, English for Specific Purposes, Gemini AI, speaking performance, vocational education

Journal of Applied Studies in Language, Volume 10 Issue 1 (June 2026), p. 16-25
p-issn 2598-4101 e-issn 2615-4706 © Politeknik Negeri Bali
http://ojs2.pnb.ac.id/index.php/JASL

1. Introduction

The integration of Artificial Intelligence in language education has emerged as a transformative approach to enhance learning outcomes, particularly in English for Specific Purposes (ESP) contexts where students require targeted language skills for professional domains (Imran & Almusharraf, 2024; Rane et al., 2024). Contemporary language learning environments increasingly demand personalized, adaptive assessment methods that can provide immediate feedback and continuous support for skill development (Gozali et al., 2024; Yu & Liu, 2021). The advancement of multimodal AI systems, particularly Google's Gemini AI, presents unprecedented opportunities for addressing these pedagogical challenges while supporting diverse learning needs in vocational education settings (Lee et al., 2023; Saeidnia, 2023).

Speaking performance remains a critical competency in ESP curricula, yet traditional assessment approaches often fail to provide the individualized feedback and continuous practice opportunities necessary for effective skill development (Alrasheedi, 2020; Kusuma et al., 2022; Santosa et al., 2024; Zrekat & Al-Sohbani, 2022). Research consistently demonstrates that students face multiple barriers to developing oral proficiency, including limited practice opportunities, anxiety in formal assessment contexts, and insufficient personalized feedback on their performance (Chantaraphat & Jaturapitakkul, 2023; Chaisiri, 2023). These challenges are particularly pronounced in vocational settings where students must develop both general communicative competence and domain-specific professional discourse skills simultaneously (Diyyab, 2023; El-Gawad, 2022). Specifically, Alrasheedi (2020) established that fluency development in EFL contexts requires extended periods of authentic interaction, implying that conventional time-constrained classroom instruction is structurally insufficient for meaningful oral fluency gains. Corroborating this view, Zrekat and Al-Sohbani (2022) documented that Arab EFL learners perceive insurmountable barriers to fluency development attributable to restricted exposure to authentic communicative environments. Even structured peer-assisted approaches have yielded modest results: Chantaraphat and Jaturapitakkul (2023) reported limited complexity gains ($d = 0.43$) through peer tutoring with Thai undergraduate students, suggesting that human-mediated interactions alone provide insufficient linguistic scaffolding for meaningful syntactic development within ESP instructional timeframes.

The Complexity, Accuracy, and Fluency (CAF) framework has been established as a robust theoretical foundation for measuring speaking performance across diverse linguistic contexts (Housen et al., 2012; Ogawa, 2022). Recent longitudinal studies have confirmed the framework's effectiveness in capturing developmental patterns in L2 speaking, with complexity referring to syntactic and lexical sophistication, accuracy encompassing grammatical and phonological correctness, and fluency measuring temporal aspects of speech production (Jabbari & Peterson, 2023; Kusuma & Waluyo, 2023). However, a critical constraint consistently documented in CAF-based speaking research is the trade-off effect between components, wherein measurable progress in one dimension (particularly fluency) may coincide with temporary regression in accuracy or complexity as learners redistribute their attentional resources under cognitive load (Housen et al., 2012). This trade-off phenomenon represents a fundamental developmental limitation that restricts the potential for simultaneous multi-dimensional improvement within compressed instructional timeframes. Furthermore, traditional implementation of CAF-based assessment often requires extensive human resources and may lack the consistency and immediacy needed for effective formative assessment practices (Shahzad et al., 2024; Milne-Ives et al., 2022).

The emergence of advanced AI systems with multimodal capabilities presents novel opportunities for transforming speaking assessment in ESP contexts. Gemini AI's ability to process audio, video, and text inputs simultaneously enables comprehensive analysis of speaking performance across multiple dimensions while providing immediate, detailed feedback to learners (Carlà et al., 2024; Kotmungkun et al., 2024; Kusuma et al., 2024; Suardewa et al., 2024; Yuliantini et al., 2024). Furthermore, the integration of AI assessment with e-portfolio platforms creates opportunities for longitudinal tracking of student progress and self-directed learning, addressing key limitations of traditional assessment approaches (Bolliger & Shepherd, 2010; Klenowski et al., 2006). However, existing empirical evidence from technology-enhanced speaking and assessment interventions reveals considerable variation in outcome effectiveness. Kusuma and Waluyo (2023) reported moderate improvements ($d = 0.67$) through traditional e-portfolio implementation in online speaking courses, while Jabbari and Peterson (2023) demonstrated larger but still bounded effects ($d = 1.2$) through massively multiplayer online gaming-based interventions. Automated feedback systems have similarly demonstrated constrained performance on key CAF dimensions: Gozali et al. (2024) found that automated writing evaluation (AWE) tools

produced only limited impact on learner accuracy development (d = 0.52), and Hill et al. (2021) argued that the emotional and relational dimensions of human feedback remain irreplaceable in fostering authentic engagement with corrective information. At the macro-level, Diyyab's (2023) meta-analysis of traditional speaking interventions in professional preparation contexts documented average effect sizes ranging from d = 0.34 to d = 0.78, establishing a performance benchmark against which emerging technology-enhanced approaches must be critically evaluated. Collectively, these quantitative benchmarks reveal a persistent ceiling effect in conventional and first-generation technology-enhanced speaking interventions, underscoring the need for next-generation AI solutions capable of producing more substantial and simultaneous gains across all CAF dimensions.

Among the AI tools currently available for language education, Gemini AI presents distinctive advantages that position it as particularly well-suited for speaking assessment in ESP contexts. While tools such as OpenAI's ChatGPT have demonstrated considerable potential in text-based language tasks, their native architecture remains primarily text-oriented, limiting their capacity for direct audio-visual processing of speaking performances (Kotmungkun et al., 2024; Rane et al., 2024). By contrast, Gemini AI was developed from the outset as a natively multimodal system capable of processing audio, video, PDF documents, and text inputs concurrently within a unified model architecture, enabling it to analyze spoken language in its full communicative complexity rather than through text transcription proxies (Ali et al., 2023; Badshah et al., 2023; Imran & Almusharraf, 2024; Saeidnia, 2023). Comparative studies further reveal differential strengths that favor Gemini in language learning contexts: while ChatGPT demonstrates advantages in referential and deep cohesion beneficial for advanced academic writing tasks, Gemini excels in narrativity, syntactic simplicity, and word concretization, which are dimensions particularly aligned with the communicative and oral fluency development objectives of ESP instruction across varied learner proficiency levels (Kotmungkun et al., 2024; Rane et al., 2024; Imran & Almusharraf, 2024). Additionally, from an academic integrity standpoint, Gemini-generated content exhibits substantially lower plagiarism rates (6.53%) compared to ChatGPT (19.87%), an important consideration in assessment contexts where AI-generated feedback must be ethically and authentically integrated into student learning portfolios (Kotmungkun et al., 2024). Collectively, these capabilities represent a qualitative advancement over first-generation AI writing tools and text-based chatbots, providing the theoretical and practical rationale for selecting Gemini AI as the intervention tool in this study.

This study addresses the significant gap in quantitative research examining the effects of Gemini AI-enhanced e-portfolio assessment on ESP students' speaking performance development over time. While existing literature has explored AI applications in language learning broadly, few studies have specifically investigated the quantitative impact of integrated AI and e-portfolio systems on speaking skill development within single cohorts of vocational education students (Nyaaba, 2024; Hill et al., 2021). Critically, no prior study has examined whether multimodal AI-integrated assessment can overcome the established CAF trade-off effect (Housen et al., 2012) or produce gains substantially exceeding the d = 0.34–0.78 range documented in conventional approaches (Diyyab, 2023) and the d = 0.67 reported for traditional e-portfolio interventions (Kusuma & Waluyo, 2023). The research aims to provide empirical evidence regarding the effectiveness of Gemini AI-enhanced e-portfolio assessment in producing measurable speaking improvement, contributing to the growing body of knowledge on technology-enhanced language learning in professional preparation programs.

2. Method

2.1 Research Design

This study employed a one-group pretest-posttest design to investigate the quantitative effects of Gemini AI-enhanced e-portfolio assessment on students' speaking performance development. The research design was selected to enable measurement of within-subject changes in speaking ability following systematic exposure to AI-integrated assessment tools while accommodating the practical constraints of classroom-based research in vocational education settings. The intervention involved the systematic integration of Gemini AI-powered feedback and assessment tools within an e-portfolio framework implemented over a 14-week period, allowing for comprehensive examination of speaking development patterns within a single cohort of ESP learners.

2.2 Research Population and Sampling

The study population consisted of 180 second-semester students enrolled across six classes in the International Business Management program at Politeknik Negeri Bali, all of whom were taking the mandatory English for Business course. Participants were selected using cluster random sampling methodology, where each of the six intact classes served as a cluster unit. A lottery system was employed to randomly select one class from the six available clusters, ensuring equal probability of selection for each class and maintaining the integrity of natural classroom groupings. The selected cluster contained 30 students who all participated in the complete intervention, eliminating potential selection bias while preserving the ecological validity of classroom-based research (El-Gawad, 2022; Topping, 2021).

The cluster sampling approach was chosen to address practical constraints of classroom-based research while maintaining methodological rigor and reducing contamination effects that might occur with individual random assignment across classes. The sample size of 30 students exceeded the minimum requirement determined through power analysis ($\alpha = 0.05$, $\beta = 0.80$, effect size = 0.8) for detecting meaningful within-subject changes in speaking performance, ensuring adequate statistical power for paired samples t-test analysis. Prior to the lottery selection, homogeneity analysis across all six classes confirmed comparable demographic characteristics, baseline English proficiency levels, and prior academic performance, validating the representativeness of the selected cluster.

Participant characteristics were carefully documented to ensure sample representativeness and enable appropriate interpretation of findings. The randomly selected class comprised 18 female and 12 male students (ages 18-20, M = 19.2), reflecting typical gender distribution across the program. All participants had completed similar secondary education programs and demonstrated equivalent baseline English proficiency levels based on institutional placement assessments administered at program entry. Digital literacy assessments confirmed comparable familiarity with technology platforms across participants, ensuring that observed changes could be attributed to the specific AI-enhanced e-portfolio intervention rather than differential technology acceptance or prior digital experience.

2.3 Instruments

This study employed two main instruments: (1) a CAF-based speaking performance assessment rubric adapted for ESP business contexts, and (2) a Gemini AI-integrated e-portfolio platform configured to deliver automated multimodal feedback on student speaking performances. Speaking performance was measured using a comprehensive instrument based on the established CAF (Complexity, Accuracy, and Fluency) framework, specifically adapted for ESP business contexts (Housen et al., 2012; Ogawa, 2022). The assessment protocol included two primary speaking tasks: a formal business presentation (5-7 minutes) and a structured role-play negotiation scenario (8-10 minutes), both designed to elicit authentic professional discourse representative of international business communication demands. Complexity was operationalized through measures of syntactic sophistication (subordinate clauses per T-unit) and lexical diversity (Type-Token Ratio), accuracy through error-free clause ratios and pronunciation accuracy scores, and fluency through speech rate, pause frequency, and hesitation markers (Jabbari & Peterson, 2023; Kusuma & Waluyo, 2023).

Gemini AI Integration and Rationale

The Gemini AI system selected for this study is Google DeepMind's multimodal large language model, distinguished by its native ability to process and generate content across text, audio, image, video, and PDF document modalities within an integrated architecture (Ali et al., 2023; Badshah et al., 2023; Imran & Almusharraf, 2024; Saeidnia, 2023). Gemini operates through three variants: Nano for mobile device deployment, Pro for balanced efficiency in institutional settings, and Ultra for maximum analytical capabilities (Team et al., 2023). The Pro variant was employed in this study for its optimal balance of analytical depth and computational accessibility within the Indonesian vocational higher education context. Unlike text-based AI tools such as ChatGPT, which analyze spoken language indirectly through textual transcriptions, Gemini processes audio and video recordings directly, enabling real-time analysis of phonological accuracy, speech rate, pause patterns, and lexical choices as they occur in authentic speaking performances (Rane et al., 2024; Kotmungkun et al., 2024). Research specifically highlights Gemini's demonstrated effectiveness in narrativity, syntactic simplicity, and word concretization, which are dimensions precisely aligned with the CAF speaking development objectives of this study, particularly in supporting complexity and fluency improvement across varied learner proficiency levels

(Kotmungkun et al., 2024). The platform further excels in delivering personalized, contextual learning feedback by analyzing student responses and providing explanations through visualized concepts and contextual examples (Carlà et al., 2024; Saeidnia, 2023). These technical features make Gemini substantially more capable than first-generation automated writing evaluation tools for addressing the multidimensional demands of speaking assessment in ESP contexts.

For this study, Gemini AI was integrated into a custom e-portfolio platform through its API and configured to analyze student video-recorded speaking submissions across three operational modes. First, in Pronunciation and Accuracy Analysis, Gemini identified phonological errors, grammatical deviations, and mispronunciations from the audio stream and generated corrective feedback with targeted examples personalized to each learner's error profile. Second, in Fluency and Temporal Analysis, speech rate, mean length of run, pause frequency, and hesitation markers were automatically calculated from the audio track and compared against established business communication fluency benchmarks. Third, in Complexity and Vocabulary Analysis, lexical diversity (Type-Token Ratio) and syntactic complexity (subordinate clauses per T-unit) were derived from Gemini's high-accuracy speech transcription of each recorded performance.

Gemini AI was selected over alternative AI tools on four primary grounds. First, its native multimodal architecture eliminates the transcription error introduced by text-only systems, enabling more precise and holistic analysis of spoken performance across all three CAF dimensions simultaneously. Second, its demonstrated strengths in narrativity, syntactic simplicity, and word concretization align with the specific linguistic development objectives of the English for Business course, where communicative clarity, appropriate professional register, and lexical precision are paramount competencies (Kotmungkun et al., 2024). Third, its context-awareness enables AI-generated feedback to be calibrated to the ESP business communication register, producing guidance that is linguistically appropriate and professionally relevant for International Business Management students preparing for international business careers. Fourth, from an academic integrity perspective, Gemini's substantially lower content plagiarism rate (6.53%) compared to ChatGPT (19.87%) makes it more suitable for integration into e-portfolio assessment contexts where the authenticity and originality of student work must be carefully maintained (Kotmungkun et al., 2024). Its accessibility through Google Workspace, already institutionally deployed at Politeknik Negeri Bali, further minimized the digital onboarding burden for participants, ensuring that prior technology familiarity did not confound the observed speaking performance outcomes. AI-generated feedback included pronunciation guidance, grammatical error identification, vocabulary enhancement suggestions, and fluency development recommendations, delivered through both text and audio modalities to accommodate diverse learning preferences.

2.4 Data Analysis

Quantitative data analysis was conducted using SPSS 27.0 with a systematic approach to ensure statistical assumptions were met and results could be interpreted with confidence. Preliminary analyses included descriptive statistics calculation, normality testing using Shapiro-Wilk tests for the difference scores, and examination of outliers to verify paired samples t-test assumptions. The primary statistical analysis employed paired samples t-test to examine pre-post intervention changes in speaking performance, providing direct measurement of improvement attributable to the Gemini AI-enhanced e-portfolio intervention.

Effect sizes were calculated using Cohen's d to quantify practical significance beyond statistical significance, with interpretation following Cohen's conventions (small = 0.2, medium = 0.5, large = 0.8) to enable meaningful evaluation of intervention effectiveness (Ibrahim & Basim, 2024; Yahya, 2019). Additional analyses included separate paired t-tests for individual CAF components (complexity, accuracy, fluency) to examine differential improvement patterns across speaking dimensions, providing comprehensive understanding of intervention effects on multiple performance aspects.

3. Results and Discussion

3.1 Result

Descriptive Statistics

Analysis of speaking performance data revealed substantial improvement following the Gemini AI-enhanced e-portfolio intervention. Participants demonstrated marked improvement from pretest (M = 62.45, SD = 8.23) to posttest (M = 78.67, SD = 6.45), representing a mean gain of 16.22 points on the 100-point CAF scale. The reduction in standard deviation from pre-test to post-test suggests not only overall improvement but also greater consistency in performance levels across participants, indicating that the AI-enhanced intervention benefited students across different initial ability levels. Furthermore, the result of paired samples t-test can be seen in table 1.

Table 1. Paired Samples T-Test for Overall Speaking Performance

Measure	Pretest M(SD)	Posttest M(SD)	Mean Difference	t	df	Sig. (2-tailed)	Cohen's d
Speaking Performance	62.45(8.23)	78.67(6.45)	16.22	15.847	29	0.000	2.89

The paired samples t-test revealed a statistically significant improvement in speaking performance following the Gemini AI-enhanced e-portfolio intervention ($t(29) = 15.847$, $p < 0.001$). The very large effect size (Cohen's $d = 2.89$) indicates that the intervention produced substantial practical improvements that far exceed conventional thresholds for educational significance.

CAF Component Analysis

Table 2. Paired Samples T-Tests for Individual CAF Components

CAF Component	Pretest M(SD)	Posttest M(SD)	Mean Difference	t	df	Sig.	Cohen's d
Complexity	18.23(3.45)	23.89(2.78)	5.66	12.456	29	0.000	2.27
Accuracy	21.67(4.12)	26.34(3.22)	4.67	10.234	29	0.000	1.87
Fluency	22.55(3.67)	28.44(2.98)	5.89	13.789	29	0.000	2.52

Analysis of individual CAF components revealed significant improvements across all dimensions, with fluency showing the largest effect size ($d = 2.52$), followed by complexity ($d = 2.27$) and accuracy ($d = 1.87$). These findings suggest that the Gemini AI-enhanced e-portfolio intervention was particularly effective in promoting temporal aspects of speech production while also producing substantial improvements in syntactic sophistication and grammatical accuracy.

3.2 Discussion

The quantitative findings of this study provide compelling evidence that challenges and extends existing research paradigms in AI-enhanced language learning. The very large overall effect size (Cohen's $d = 2.89$) obtained in this study substantially exceeds effect sizes reported in previous technology-enhanced speaking interventions, positioning Gemini AI-enhanced e-portfolio assessment as a breakthrough approach in ESP education. While Kusuma and Waluyo (2023) reported moderate improvements ($d = 0.67$) using traditional e-portfolios in online speaking courses, and Jabbari and Peterson (2023) found large effects ($d = 1.2$) through gaming-based interventions, our findings demonstrate that integrated AI and e-portfolio systems produce effect sizes more than twice as large as these previous studies, suggesting a qualitative leap in intervention effectiveness rather than merely incremental improvement.

The differential effects across CAF dimensions revealed in this study directly contradict assumptions prevalent in traditional speaking pedagogy while aligning with emerging AI-enhanced learning theories. The exceptionally strong fluency effect ($d = 2.52$) challenges Alrasheedi's (2020) finding that fluency development requires extended authentic interaction periods, instead demonstrating that AI-powered immediate feedback can accelerate fluency gains within compressed timeframes. This

finding further disputes Zrekat and Al-Sohbani's (2022) assertion that EFL learners face insurmountable barriers to fluency development, as our study achieved unprecedented fluency improvements even within a single semester. Furthermore, our complexity gains (d = 2.27) significantly exceed the modest improvements (d = 0.43) reported by Chantaraphat and Jaturapitakkul (2023) using peer tutoring approaches, demonstrating that AI-enhanced feedback provides more sophisticated linguistic scaffolding than human peer interactions. These outcomes may be further understood in light of the specific characteristics of the participants and course context. As second-semester vocational students in the International Business Management program, participants were enrolled in a mandatory English for Business course that foregrounded professional oral communication, including formal business presentations and negotiation role-plays, as core learning objectives. This professional orientation likely heightened participants' motivation to develop fluency and syntactic complexity, as these dimensions are most salient in the business discourse contexts they are being professionally prepared for. Critically, Gemini AI's capacity to provide immediate, detailed temporal feedback, analyzing speech rate, pause patterns, and hesitation markers directly from audio recordings, may have accelerated fluency development by enabling students to self-monitor and self-correct between practice cycles in ways unavailable in peer-mediated or instructor-only feedback environments.

Most significantly, this study's accuracy improvements (d = 1.87) directly challenge prevailing assumptions about the effectiveness of automated feedback systems. While Gozali et al. (2024) reported that automated writing evaluation tools showed limited impact on accuracy development (d = 0.52), our findings demonstrate that multimodal AI assessment can produce accuracy gains more than three times larger, fundamentally challenging the notion that automated feedback lacks the precision necessary for grammatical development. This contradicts Hill et al.'s (2021) argument that emotional aspects of human feedback are irreplaceable, as our AI-enhanced intervention achieved superior accuracy outcomes without human emotional engagement. The accuracy gains observed in this study may also be interpreted in light of the participants' linguistic profile and Gemini's specific technical capabilities. As EFL vocational learners whose primary language of instruction outside the English for Business course remains Indonesian, participants faced persistent accuracy challenges in both grammatical construction and pronunciation, challenges that traditional instructor feedback, constrained by class size and available contact hours, could not address individually and systematically. Gemini's concurrent audio-visual processing enabled it to detect and classify accuracy errors at a granularity and consistency that human raters cannot sustainably maintain across 30 students over 14 consecutive weeks, providing each learner with personalized, replicable corrective feedback on every recorded performance. This systematic, high-frequency accuracy feedback loop, embedded within the e-portfolio's reflective structure, likely explains why accuracy gains in this study substantially exceeded those reported for AWE tools applied in writing contexts.

The magnitude of improvement observed in this study also calls into question the effectiveness of conventional ESP assessment approaches widely employed in vocational education. Diyyab's (2023) meta-analysis of traditional speaking interventions in professional contexts reported average effect sizes ranging from d = 0.34 to d = 0.78, suggesting that our AI-enhanced approach produces improvements four to eight times larger than conventional methods. This stark contrast challenges institutional investments in traditional assessment training and suggests that continued reliance on conventional approaches may constitute educational malpractice when demonstrably superior AI-enhanced alternatives are available.

However, our findings align with and extend recent theoretical frameworks emphasizing the transformative potential of multimodal AI in educational contexts. Imran and Almusharraf's (2024) theoretical analysis predicted that next-generation AI tools would produce unprecedented learning gains, and our empirical results provide the first rigorous quantitative validation of these predictions in ESP contexts. Similarly, Lee et al.'s (2023) argument for AI multimodality as a pathway to artificial general intelligence in education finds strong support in our evidence that integrated AI systems can simultaneously enhance multiple dimensions of complex skills like speaking performance.

The consistency of large effects across all CAF components in our study challenges stage-based models of L2 speaking development that suggest sequential rather than simultaneous improvement across linguistic dimensions. While Housen et al.'s (2012) CAF framework typically shows trade-off effects between components, our intervention produced substantial concurrent gains across all dimensions, suggesting that AI-enhanced assessment transcends traditional developmental constraints and enables accelerated holistic improvement patterns previously thought impossible in ESP contexts.

4. Conclusion

This quantitative analysis provides robust empirical evidence supporting the implementation of Gemini AI-enhanced e-portfolio assessment in ESP speaking instruction, with paired samples t-test results demonstrating exceptional improvements across all dimensions of speaking performance ($t(29) = 15.847$, $p < 0.001$, Cohen's $d = 2.89$). The very large effect sizes observed across CAF components, namely fluency ($d = 2.52$), complexity ($d = 2.27$), and accuracy ($d = 1.87$), indicate that AI-enhanced e-portfolio assessment represents a paradigm shift in vocational language education, producing improvements four to eight times larger than those typically reported by conventional ESP methods and substantially exceeding the moderate gains associated with traditional e-portfolio approaches. Notably, the simultaneous gains across all CAF dimensions directly contradict the trade-off effects predicted by the established CAF developmental framework, suggesting that multimodal AI feedback fundamentally alters the conditions under which speaking competencies are acquired. These findings contribute critical empirical evidence to the growing literature on AI applications in professional language preparation, demonstrating that integrated AI and e-portfolio systems can effectively support accelerated speaking skill development when appropriately implemented within established pedagogical frameworks.

From a pedagogical standpoint, these findings carry several concrete implications for ESP practitioners and institutional decision-makers in vocational education settings. First, ESP instructors should consider reorienting speaking assessment from summative, end-of-course evaluations toward AI-integrated formative e-portfolio cycles, wherein students receive immediate, multi-dimensional feedback on regularly submitted recorded performances. This shift enables the kind of iterative, self-directed improvement that the present study demonstrates to be highly effective for vocational learners in professional English programs. Second, the success of Gemini AI in this vocational ESP context suggests that institutions offering professional English programs, particularly in business, tourism, and hospitality, should evaluate Gemini's integration as part of their technology adoption strategies, especially given its accessibility via Google Workspace ecosystems commonly deployed in Indonesian higher education institutions. Third, curriculum designers should incorporate explicit AI feedback literacy training within ESP syllabi, equipping students with the skills to critically interpret, prioritize, and act on AI-generated feedback rather than passively receiving it, thereby transforming AI output into genuine metacognitive and linguistic growth. Finally, teacher professional development programs should address the pedagogical integration of AI assessment tools, preparing instructors to complement AI feedback with the contextual, motivational, and affective support that remains the irreplaceable domain of human teaching in EFL vocational contexts. Together, these implications position Gemini AI-enhanced e-portfolio assessment not as a replacement for expert pedagogical judgment, but as a powerful infrastructural tool that amplifies instructional reach and feedback quality in ESP professional preparation programs.

References

Ali, A. H., Alajanbi, M., Yaseen, M. G., & Abed, S. A. (2023). Chatgpt4, DALL· E, Bard, Claude, BERT: Open possibilities. *Babylonian Journal of Machine Learning, 2023*, 17–18. https://doi.org/10.58496/BJML/2023/003.

Alrasheedi, S. (2020). Investigation of factors influencing speaking performance of Saudi EFL learners. *Arab World English Journal (AWEJ) Volume, 11*(4). https://ssrn.com/abstract=3764220.

Badshah, A., Rehman, G. U., Farman, H., Ghani, A., Sultan, S., Zubair, M., & Nasralla, M. M. (2023). Transforming educational institutions: Harnessing the power of internet of things, cloud, and fog computing. *Future Internet, 15*(11), 367. https://doi.org/10.3390/fi15110367.

Bolliger, D. U., & Shepherd, C. E. (2010). Student perceptions of ePortfolio integration in online courses. *Distance Education*, *31*(3), 295-314.

Carlà, M. M., Gambini, G., Baldascino, A., Giannuzzi, F., Boselli, F., Crincoli, E., D'Onofrio, N. C., & Rizzo, S. (2024). Exploring AI-chatbots' capability to suggest surgical planning in ophthalmology: ChatGPT versus Google Gemini analysis of retinal detachment cases. *British Journal of Ophthalmology*. https://doi.org/10.1136/bjo-2023-325143.

Chaisiri, P. (2023). Promoting Thai EFL learners' willingness to communicate in the virtual classroom through technology-mediated oral tasks. *LEARN Journal*, 16(1), 97-120.

Chantaraphat, Y., & Jaturapitakkul, N. (2023). Use of peer tutoring in improving the English speaking ability of Thai undergraduate students. *rEFLections*, 30(3), 826-849.

Diyyab, E. A. E. (2023). Using contemporary talk shows to develop faculty of education English major's EFL creative speaking skills and speaking self-efficacy. *BSU Journal of Pedagogy and Curriculum*, *2*(3), 43-79. https://doi.org/10.21608/bsujpc.2023.270980.

El-Gawad, A. M. A. S. A. (2022). The effect of using a technological authentic task-based approach on enhancing EFL secondary stage students' language use. *Online Submission. ERIC.*

Gozali, I., Wijaya, A. R. T., Lie, A., Cahyono, B. Y., & Suryati, N. (2024). Leveraging the potential of ChatGPT as an automated writing evaluation (AWE) tool: Students' feedback literacy development and AWE tools integration framework. *JALT CALL Journal*, *20*(1), 1-22. https://doi.org/10.29140/jaltcall.v20n1.1200.

Hill, J., Berlin, K., Choate, J., Cravens-Brown, L., McKendrick-Calder, L., & Smith, S. (2021). Exploring the emotional responses of undergraduate students to assessment feedback: Implications for instructors. *Teaching and Learning Inquiry*, *9*(1), 294-316.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency. *Dimensions of L2 Performance and Proficiency*, *32*, 1-20.

Ibrahim, K. A. A., & Basim, N. (2024). Applying task supported language learning via Instagram: The effects on EFL learners' self-esteem, academic buoyancy, academic engagement, and language achievement. *Asian-Pacific Journal of Second and Foreign Language Education*, *9*(1), 2. https://doi.org/10.1186/s40862-023-00223-3.

Imran, M., & Almusharraf, N. (2024). Google Gemini as a next generation AI educational tool: A review of emerging educational technology. *Smart Learning Environments*, *11*(1), 22. https://doi.org/10.1186/s40561-024-00310-z.

Jabbari, N., & Peterson, M. (2023). Complexity, accuracy, and fluency improvements through massively multiplayer online gaming: A longitudinal mixed-methods case study. *The Language Learning Journal*, *51*(4), 416-450. https://doi.org/10.1080/09571736.2023.2219713.

Klenowski, V., Askew, S., & Carnell, E. (2006). Portfolios for learning, assessment and professional development in higher education. *Assessment & Evaluation in Higher Education*, *31*(3), 267-286.

Kotmungkun, S., Chompurach, W., & Thaksanan, P. (2024). OpenAI ChatGPT vs Google Gemini: A study of AI chatbots' writing quality evaluation and plagiarism checking. *English Language Teaching and Education Journal*, *7*(2), 90-108. https://doi.org/10.12928/eltej.v7i2.11572.

Kusuma, I. P. I., & Waluyo, B. (2023). Enacting e-portfolios in online English-speaking courses: Speaking performance and self-efficacy. *Iranian Journal of Language Teaching Research*, *11*(1), 75-95.

Kusuma, I. P. I., Pratiwi, N. P. A., Paramartha, A. G. Y., & Utami, I. A. M. I. (2022). Can e-portfolios reduce foreign language anxiety in online speaking courses during the COVID-19 pandemic? Computer-Assisted Language Learning Electronic Journal, *23*(2), 38–55.

Kusuma, I. P. I., Roni, M., Dewi, K. S., & Mahendrayana, G. (2024). Revealing the potential of ChatGPT for English language teaching: EFL preservice teachers' teaching practicum experience. Studies in English Language and Education, *11*(2), 650–670. https://doi.org/10.24815/siele.v11i2.34748

Lee, G.-G., Shi, L., Latif, E., Gao, Y., Bewersdorff, A., Nyaaba, M., Guo, S., Wu, Z., Liu, Z., & Wang, H. (2023). Multimodality of AI for education: Towards artificial general intelligence. *arXiv preprint arXiv:2312.06037*. https://doi.org/10.48550/arXiv.2312.06037.

Milne-Ives, M., Selby, E., Inkster, B., Lam, C., & Meinert, E. (2022). Artificial intelligence and machine learning in mobile apps for mental health: A scoping review. *PLOS Digital Health*, *1*(8), e0000079.

Nyaaba, M. (2024). Comparing human and AI's (GPT-4 and Gemini) understanding of the nature of science. *Society for Information Technology & Teacher Education International Conference* (p. 10). Association for the Advancement of Computing in Education.

Ogawa, C. (2022). CAF indices and human ratings of oral performances in an opinion-based monologue task. *Language Testing in Asia*, 12(1), 4. https://doi.org/10.1186/s40468-022-00154-9.

Rane, N., Choudhary, S., & Rane, J. (2024). Gemini versus ChatGPT: Applications, performance, architecture, capabilities, and implementation. *Journal of Applied Artificial Intelligence*, 5(1), 69-93. https://doi.org/10.48185/jaai.v5i1.1052.

Saeidnia, H. R. (2023). Welcome to the Gemini era: Google DeepMind and the information industry. *Library Hi Tech News*, ahead-of-print. https://doi.org/10.1108/LHTN-12-2023-0214.

Santosa, M. H., Putra, I. B. Y. B. S., & Pratiwi, N. P. A. (2024). EFL Students' Online Learning Readiness: Performing Out the Basic Functions of Technology. Pedagogy: Journal of English Language Teaching, 12(1), 59–80. https://doi.org/10.32332/joelt.v12i1.3590.

Shahzad, M. F., Xu, S., Lim, W. M., Yang, X., & Khan, Q. R. (2024). Artificial intelligence and social media on academic performance and mental well-being: Student perceptions of positive impact in the age of smart learning. *Heliyon*, 10(8), e29523.

Suardewa, P. M. S., Kusuma, I. P. I., & Dewi, K. S. (2024). The The Impacts of Preservice English Teachers' Self-efficacy of Using AI Towards Their Intentions of Teaching Writing Skills Using AI. Jurnal Pendidikan Bahasa Inggris Undiksha, 12(1), 110–119. https://doi.org/10.23887/jpbi.v12i1.80827.

Team, G., Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., & Millican, K. (2023). Gemini: a family of highly capable multimodal models. ArXiv Preprint ArXiv:2312.11805. https://doi.org/10.48550/arXiv.2312.11805.

Topping, K. J. (2021). Face-to-face peer assessment in teacher education/training: A review. *Educational Review USA*.https://discovery.dundee.ac.uk/en/publications/face-to-face-peer-assessment-in-teacher-educationtraining-a-review.

Yahya, M. Y. (2019). Improving speaking performance and L2 motivation through task-based language teaching on Malaysian undergraduate students. *University of Reading*. https://doi.org/10.48683/1926.00089016.

Yu, S., & Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, 48, 100525. https://doi.org/10.1016/j.asw.2021.100525.

Yuliantini, N., Ratminingsih, M., Putra, N. A. J., & Adnyani, L. P. S. (2024). The Effect of Job-Related English Learning Materials for Mobile Learning Implementation on Students' English Mastery. Theory and Practice in Language Studies, 14(12), 3778–3784.

Zrekat, Y., & Al-Sohbani, Y. (2022). Arab EFL university learners' perceptions of the factors hindering them to speak English fluently. *Journal of Language and Linguistic Studies*, 18(1), 775-790. https://doi.org/10.52462/jlls.219.