Journal of Linguistics, Literature, and Culture

Volume 8, Number 2, November 2026
p. 241-256
http://journal2.uad.ac.id/index.php/notion/index

AI benefits, challenges, and pedagogical implications on speaking task completion: A multiple case study of two vocational programs in Indonesia

Arini Sabrina[1]*, Ajar Pradika Ananta Tur[2]
[1]Politeknik LPP Yogyakarta, Yogyakarta, Indonesia
[2]Universitas Ahmad Dahlan, Yogyakarta, Indonesia

*Corresponding author: arini@polteklpp.ac.id

Citation: Sabrina, A. & Tur, A.P.A. (2026). AI benefits, challenges, and pedagogical implications on speaking task completion: A multiple case study of two vocational programs in Indonesia. *NOTION: Journal of Linguistics, Literature, and Culture, 8*(*2*), 241-256. https://doi.org/10.12928/notion.v8i2.14271

ARTICLE INFO

Article History:
Received: 2025-08-15
Accepted: 2026-09-27

Keywords:
Artificial intelligence (AI)
English for specific purpose
Speaking skills
Vocational context
Politeknik LPP Yogyakarta

This work is licensed under a Creative Commons Attribution 4.0 International License

ABSTRACT

This study investigates spoken English performance in vocational education contexts. Conducted at Politeknik LPP Yogyakarta, a plantation-focused vocational institution, the research adopts a multiple case study design with an embedded quantitative component involving two distinct student groups: 30 Accountancy students and 28 Plantation Crop Farming students. All data were collected with voluntary, informed participation and reported using anonymized participant codes. Findings show that AI use supported time efficiency, task completion, material comprehension, and communicative confidence across both groups. At the same time, the participants showed overreliance on AI-generated language, shallow comprehension of AI-generated content, limited preparation for adapting and internalizing AI-assisted material, and occasional mismatches between AI output and task requirements, leading to the AKT cohort's AI-assisted mean (72.93) falling below its own prior non-AI mean (79.67) as its paradox to the 'material comprehension' as AI benefit aforementioned. These findings suggest that AI functions best as a supportive tool rather than a substitute for learner preparation and instructor guidance, particularly in technical vocational fields where accuracy of field-specific vocabulary matters. For the pedagogical implication more broadly, the study highlights the need for structured guidance on AI use alongside speaking skill development. Educators are encouraged to pair AI tools with explicit structured discourse-building and motivation-building strategies. Future research should examine a broader range of AI tools, larger samples, and longer-term effects on spoken language development.

I. INTRODUCTION

Indonesia's plantation sector remains a major economic driver, and its integration into international commodity markets means Diploma III graduates in plantation-related fields increasingly need English communication skill alongside their technical expertise. That need reframes English, at the vocational level, as a functional workplace competence rather than a general academic subject. Two constraints work against that aim in practice: limited instructional hours, and students' sense that English is peripheral to their chosen major (Marwan, 2009). It is this second constraint, more than the first, that has pushed institutions toward supplementary strategies, among which the integration of Artificial Intelligence (AI) has drawn particular attention.

AI itself is not new as the field dates to 1956 (Haenlein & Kaplan, 2019), but its recent classroom presence is. Tools such as ChatGPT, QuillBot, Perplexity, and Grammarly, broadly understood as technology approximating human cognitive functions (Sheikh, Prins, & Schrijvers, 2023), are now used routinely to help the language learners in completing language tasks such as creating poetries or short stories (Ningrum et al, 2026). Why this might matter for motivation specifically has a plausible mechanism behind it: Patall, Cooper, and Robinson (2008) found that extended learning duration can itself become demotivating, which suggests that whatever efficiency AI offers may carry a motivational benefit beyond simple convenience, which is consistent with Self-Determination Theory account of how reduced task burden can support autonomous engagement rather than merely relieving controlled compliance (Deci & Ryan, 2016).

Whether that benefit shows up empirically depends on how a given study frames motivation in the first place. Several treat it as a background condition for achievement rather than a direct outcome of AI use. Stander (2022) and Bećirović, Dubravac, and Brdarević-Čeljo (2022) both take this approach, while Yang and Wang (2022) fold it into a wider set of affective and socio-economic variables. Studies that isolate AI's effect on motivation more directly report a more consistent, if narrower, picture: Ali, Shamsan, Hezam, and Mohammed (2023) link ChatGPT use to increased motivation for writing and reading; Al-Raimi, Mudhsh, Muqaibal, and Al-Yafaei (2024) report reduced anxiety and improved productive skills across 2022–2023 literature; and Wei (2023), using an experimental design, finds AI-assisted instruction outperforming traditional instruction on autonomy, engagement, and motivation. Almehmadi's (2024) systematic review of thirteen studies (2019–2023) tempers this optimism usefully, concluding that AI supports language learning only when paired with complementary teaching methods, not as a standalone fix, as also stated by Sabrina (2020) who highlights the need for language educators to be adaptive and creative in the teaching process. Across all of this, the setting is almost always general English language learning. Vocational learners barely appear.

They appear even less once the literature turns to specific skill outcomes, and where it does, the coverage is lopsided. Speaking skills present a further complication, as AI assistance in this area faces more tangible difficulties within the foreign language learning context (Adelita & Daulay, 2023). Pronunciation is the best-supported case: Sariani et al. (2022) and Kazu and Kuvvetli (2023) both report AI-assisted feedback improving pronunciation accuracy and retention, since AI feedback functions as scaffolding device, a term taken from Vygotsky's Zone of Proximal Development in Wass et all (2011), that narrows the gap between a learner's current ability and their potential performance with supporta. Halimah et al. (2025) is one of few studies to carry this into a vocational context, using AI-based voice chatbots. Reading gains appear in Wang and Yan (2022), but again outside a vocational setting, while Zawacki-Richter et al. (2019) document AI support for grammar, pronunciation, vocabulary, and writing in higher education broadly, without narrowing to vocational learners at all. Only two studies push achievement findings into a genuinely specific-purpose setting — Kavgić (2024) with adult English learners at an IT company, and Xatamova and Ashurov (2024) with legal English learners. Both stop at a single professional domain, leaving vocational and plantation-adjacent education untouched.

Besides the absence of vocational and plantation-adjacency, none of the related studies are free of challenges on AI-assisted language learning. Chea and Xiao (2024) identify overreliance and ethical concerns in AI-assisted reading; Ahimsa and Khawa (2025) point to teachers' limited preparation time for integrating AI into instruction; Al-khresheh (2024) adds skill degradation and output skew to the list. These risks land harder, arguably, on vocational learners specifically. Teaching EFL in vocational settings already differs from traditional academic contexts, given its focus on industry-specific communication skills, practical language use, and immediate applicability, and it is compounded by students' lack of motivation, since English is not their main major (Marwan, 2009).

What direct evidence exists from vocational settings is thin and scattered across single-tool, single-skill studies. Purbasari (2025) finds Talkpal.AI useful for general English practice among vocational students but weak in a specific classroom. Dja'far and Hamidah (2024) report success with Google Read Along among diploma students, but confined to pronunciation. Romadhon (2025) finds chatbot-assisted vocabulary gains among polytechnic students, constrained by a small sample and one skill domain. Not one of these studies compares AI use across more than one vocational major, lets students draw on multiple AI tools as they naturally would, or examines a productive skill as complex as speaking under authentic assessment conditions.

Put together, this leaves three gaps this study is positioned to address. Vocational context-focused AI researches have stayed tool-specific rather than reflecting the range of AI applications students actually reach for when left to choose freely. They have rarely compared

cohorts from different technical majors within a single institution, leaving open whether AI's benefits and risks hold consistently across disciplinary backgrounds. Furthermore, they have given speaking performance under realistic assessment formats such as a speech and a job interview almost no qualitative attention, staying instead with isolated pronunciation drills. This study takes up all three through a qualitative descriptive multiple-case study of how AI use shaped the speech and job interview performance of Accountancy (AKT) and Plantation Crop Farming (BTP) Diploma III students at Politeknik LPP Yogyakarta.

Specifically, it asks: (1) what benefits do two groups report from using AI to prepare their speech and job interview performances; (2) what challenges emerge from AI use across the two majors and the two task types; and (3) what differences, if any, distinguish the two groups in their patterns of AI use, comprehension, and performance outcomes.

II. METHOD

This study addresses the gaps identified above through a qualitative descriptive approach using a multiple case study design (Ebneyamini & Moghadam, 2018) yet embedded with a quantitative component as supporting data. The two majors (30 Accountancy (AKT) and 28 Plantation Crop Farming (BTP) Diploma III students at Politeknik LPP Yogyakarta) were treated as two distinct cases rather than a single combined sample, since case study design calls for units that share a common context but differ on a theoretically relevant dimension. Here, both groups took the same English course, under the same lecturer-researcher, in the same institutional setting, but differed in disciplinary major and, correspondingly, in the technical vocabulary and professional register their speaking tasks required. This shared-context, single-variable-difference structure allowed within-case analysis of each major's AI use pattern before a cross-case comparison was drawn, addressing directly what makes the two groups analytically distinct "cases" rather than simply two classes. Participants were included through total sampling: all students enrolled in the two English classes took part in the classroom activities from which data were drawn, with no additional selection criteria applied.

Both groups completed two speaking tasks over the course of the study conducted across 4 weeks. The first was a persuasive speech, with topics constrained to plantation-related themes for BTP students and economic themes for AKT students to keep the task relevant to each major's professional register. The second was a simulated job interview, structured around lecturer-provided blueprint questions and model answers. For the speech task, students were given one week to prepare before performing live and without notes, assessed against four criteria: content, pronunciation, intonation, and body language. Students were free to use any AI tool during preparation; questionnaire and interview data later showed that Gemini, Copilot, ChatGPT, Blackbox, and Perplexity were the tools most frequently used, primarily for script arrangement, paraphrasing, and pronunciation checking.

Four data sources were used to enable triangulation. The questionnaire, administered to all participants after each performance task, combined closed items — asking, for example, how much preparation time students spent, whether and how much AI was used, which specific parts of the task AI assisted with (text drafting, pronunciation, or other components), and how frequently AI was used across coursework generally — with some of them were open-ended items asking students to reflect on the benefits and drawbacks they experienced using AI during and after the task, and whether they felt able to complete a similar task without AI assistance in the future. This combination allowed the questionnaire to capture both the extent of AI use and students' own evaluative reflection on it, rather than usage data alone. To minimize the risk of response bias, the questionnaire was administered only after lecturers had completed scoring of student performance, ensuring that students' self-reported responses were not influenced by prior knowledge of their evaluation outcomes. Group interviews were then conducted with all the participants from each major, implementing voluntary-response sampling to probe preparation experiences and challenges in more depth than the questionnaire format permitted. Again, this data source was the conducted after the students' performances were scored to minimize the response bias risk. Observational notes were recorded by the lecturer-researcher during both preparation and performance sessions, focusing on delivery quality, apparent comprehension of the material being presented, and behavioral indicators of AI dependency, such as hesitation when asked to elaborate beyond a scripted answer. Finally, document analysis was conducted on the participants' AI-assisted scripts and drafts to examine directly the extent and nature of AI's contribution to the final submitted text, rather than relying solely on students' self-reported estimates of AI involvement.

Data analysis followed Ebneyamini & Moghadam's stages (2018): examining, categorizing, tabulating, and drawing propositions/conclusions. Qualitative responses from the open-ended questionnaire items, interview transcripts, and observational notes were read repeatedly and coded inductively for recurring ideas; codes were then grouped into broader categories that ultimately formed the study's four benefit themes (time efficiency, project completion process, material understanding, self-confidence) and three challenge themes (overreliance, comprehension gaps, output mismatch). Quantitative descriptive data from the closed questionnaire items, reported as frequencies and percentages, were tabulated separately and used to indicate the relative weight of each qualitative theme across the two majors and the two task types, rather than to test statistical significance. Mean performance scores were calculated in the same descriptive manner to characterize overall achievement per major, although in the future research, there is a recommendation to establishing inter-rater reliability. Findings from all four sources were then triangulated to identify convergent and divergent patterns, which are presented thematically in the following section.

Because data were drawn from students enrolled in the researcher's own class, several ethical safeguards were applied. Participation in the questionnaire and interview components was voluntary and did not affect students' course grades. Participants were informed of the study's purpose prior to data collection. All questionnaire and interview responses reported in this article are anonymized using participant codes (e.g., Questionnaire Participant 1 as Q1, Interview Participant 1 as I1) rather than names.

III. RESULTS AND DISCUSSION

Roles of AI in ESP learning

Across both majors, the participants identified four recurring benefits of AI use in preparing their speech and job interview performances: time efficiency, support with the project completion process, material understanding, and self-confidence. These themes were not equally distributed across the two task types, however, nor did they carry the same weight for BTP and AKT students, a pattern that becomes clearer when the evidence for each theme is examined in turn between the comprehension gains reported here and the shallower forms of understanding reported under the challenges theme below started with the table of participant distributions to show the detailed information and aid the readers to read the analysis.

Table 1. Participants' distributions

Roles of AI in ESP learning				
No.	**Theme**	**Major**	**Task**	**Frequency**
1	Time Efficiency	AKT and BTP	Speech	58
2	Project Completion Progress		Speech and Interview	
3	Material Understanding		Speech and Interview	
4	Self-Confidence		Speech	
Challenges in AI-mediated ESP speaking performance				
No.	**Theme**	**Major**	**Task**	**Frequency**
1	Preparation Time and AI Reliance	AKT and BTP	Speech and Interview	AKT = 30 BTP = 28
2	Comprehension Paradox			
3	Diminished Language Autonomy			
4	Output Mismatch: Register and Coherence			
5	An Absent Concern			

Time efficiency

Twenty-two participants across both majors, concentrated in the speech task, described AI as primarily a time-saving tool rather than a means of improving content quality. The distinction matters: what they valued was not that AI made their work better, but that it made the work possible within the time they had. Several linked this directly to competing demands such as coursework load, illness, or approaching deadlines that left little room for language

preparation on top of everything else. One participant said AI let them finish the task early enough that they stopped worrying about the consequences of submitting late (Q1); another reported performing well despite preparing in a single day (Q2). Read together, these responses suggest AI functioned less as a shortcut around effort and more as a buffer against time scarcity specifically, a distinction the data does not yet let us extend confidently to BTP versus AKT students individually, since the 22 respondents were not broken down by major in the questionnaire summary. This pattern is consistent with Patall, Cooper, and Robinson's (2008) finding that shorter learning duration can itself be motivating, and with Al-Raimi, Mudhsh, Muqaibal, and Al-Yafaei's (2024) report of reduced anxiety among AI-assisted foreign language learners, though the present data speak to perceived time pressure rather than anxiety as a measured construct, and the comparison should be read as consistent with, not a direct replication of, either study.

Project completion process

Where the previous theme concerned time, this one concerned craft. Students here generally had adequate time to meet the deadline but described uncertainty about how to construct the work itself like script structure, pronunciation, up to grammar and used AI as a collaborator on execution rather than a substitute for their own ideas. One participant described abandoning dictionary and Google Translate use entirely after the lecturer introduced AI for the speech task, and admitted becoming reliant on it in subsequent assignments (Q3); another reported receiving minimal correction from the lecturer because AI-assisted vocabulary was largely accurate to begin with (Q4). In interview-task responses specifically, participants additionally credited AI with helping calibrate the formal register the roleplay format demanded such as pronunciation, grammar use, up to formal expressions. This suggests AI's contribution to project completion was not uniform across task types: it appears to have supported technical correctness in both, but supported register calibration specifically where the task required it. This finding is broadly consistent with Xatamova and Ashurov's (2024) work on AI-assisted legal English aforementioned.

Material understanding

Fifteen of 58 speech-task participants reported that AI offered a supplementary route to understanding course material, alongside the lecturer's explanation and the module text. Framed this way, AI functioned less as a replacement for instruction than as an additional channel suited to students whose learning preferences the standard lecture-and-module format did not fully serve. Interview-task participants echoed this, describing AI as a resource for building background knowledge; one participant specifically credited it with helping them understand the content of their own spoken interview responses (Q5). This is a notable detail, since it implies some participants were using AI not just to produce language but to comprehend what they themselves were about to say. That detail is worth holding onto: it

anticipates a tension this article returns to under the challenges theme, where AI-assisted comprehension and shallow or incomplete comprehension are reported side by side. The two findings are not necessarily contradictory a language learner can gain understanding of some material through AI while still comprehending other parts only superficially, but the coexistence deserves more direct treatment than presenting the benefit in isolation would allow, and is taken up further below.

Self-confidence

A smaller subset of speech-task students, two of 58 students, described becoming more confident and active when delivering their speech to the class. Small as this subset is, the direction of the finding aligns with Siahaan, Ardayanti, and Ramasari's (2022) account of confidence as a pathway to stronger motivation and, in turn, better achievement, and with Wei's (2023) finding that AI use can support learner autonomy and engagement. Read alongside the material understanding theme above, a plausible mechanism emerges: increased comprehension may be feeding confidence for at least some students, rather than the two operating as independent benefits. The present data cannot establish this causally since the sample here is too small and the finding too thinly reported to bear much interpretive weight on its own, but it is worth stating as a mechanism for future research to test directly, rather than treating confidence as a standalone, self-explanatory benefit.

Challenges in AI-mediated ESP speaking performance

Where the previous section addressed what AI contributed to speaking preparation, this section examines where that contribution broke down, specifically, where AI mediation introduced a gap between what students produced and what they could independently comprehend or reproduce. Four patterns emerged: compressed preparation time correlating with heavy AI reliance, a comprehension paradox in which high self-reported understanding coexisted with demonstrably shallow processing, diminished language autonomy, and register or coherence mismatches in AI output.

Preparation time and AI reliance

Despite a full week allocated for speech preparation, most participants began two to three days before performance day: 53.3% of AKT participants for the speech task and 50% for the interview task, against 60.7% (speech) and 41.4% (interview) of BTP participants. This compression tracked closely with AI dependency — 25 of 28 BTP students reported using AI for more than half of their work, compared with 12 of 30 AKT students, with several AKT students describing AI as producing their scripts entirely. In interviews, students linked the delay directly to competing demands on their time rather than to disengagement from the task itself: *"it's because of schedule compression... almost every day there is always a task that should be submitted"* (I1), and *"we are still preparing for gathering project"* (I2). Read as language-

production data rather than a general time-management complaint, this suggests AI functioned as a compensatory mechanism for compressed drafting time specifically. The tool did not create the shortfall, but it absorbed the consequence of it, and did so more heavily among BTP students than AKT students.

Comprehension paradox

Self-reported comprehension was high on its face: 66.7% (interview) and 53.3% (speech) of AKT participants, and 79.3% (interview) and 71.4% (speech) of BTP participants, described "full" understanding of their own AI-assisted content, several attributing this to their ability to integrate AI output with their own reasoning. Set against this, though, is a substantial minority reporting the opposite: 43.3% of AKT speech-task participants and 28.6% of BTP speech-task participants described their understanding as superficial or absent, with one BTP participant stating outright that they *did not understand* the AI-generated work they submitted. This is not a minor inconsistency; it is the central interpretive problem of this dataset, and it becomes sharper once performance data are brought in. AKT participants' mean speech score was 72.93, against a prior non-AI-related mean of 79.67 for the same cohort, a decline that sits uneasily next to a self-reported comprehension rate above 50%. The most defensible reading is not that the participants were being untruthful about their comprehension, but that "understanding" as self-reported here likely reflects recognition and successful reproduction of AI-generated language: being able to read, paraphrase, or deliver the content fluently rather than independent command of it as language they generated and could defend under pressure. Delivery-level evidence supports this distinction directly: several AKT participants appeared to have memorized scripted content without matching intonation or stress to meaning, and the lecturer's observational notes recorded mispronunciation and lexical items students had not encountered before using AI to produce their script. Comprehension, in other words, may be better understood here as a matter of processing depth. Yet, this general statement actually needs stronger empirical support rather than relying solely on a few participants' comments.

Diminished language autonomy

Beyond comprehension specifically, the participants described a broader erosion of independent language production. This aligns with what Chea and Xiao (2024) report as overreliance in AI-assisted reading, extended here to speaking preparation: *"the negative impact of AI is that it makes users increasingly dependent on its use, so that we become a little trivial about tasks and do not want to try hard to do tasks independently"* (Q1), and *"the results I got were less satisfactory, in my opinion, because it was not purely from my own head"* (Q2). Notably, the frequency of overreliance comments dropped from the speech task to the interview task (22 to 11 mentions), while comprehension-related complaints rose to become the dominant concern in the interview task (13 mentions) instead. This shift is worth pausing on rather than

treating as noise: the interview task came with lecturer-provided blueprint questions and model answers, giving students a fixed linguistic scaffold to work from, whereas the speech task required them to generate original content within only a topic constraint. Less linguistic freedom, in other words, appears to have reduced opportunities for the kind of unchecked AI delegation that produces overreliance complaints, but it did not reduce, and may have shifted attention toward, the deeper problem of whether students could process the language they were given. Two AKT participants who prepared their interview entirely without AI, relying solely on the lecturer's blueprint, offer an informal counterpoint here: structured linguistic input from the instructor, without AI mediation, appears to have been sufficient scaffolding on its own for at least some students. A detail with direct implications for how much AI supports a well-specified task actually requires, taken up below. The students also linked this dependency to a narrower sense of their own linguistic capacity: *"it makes me too dependent on AI assistance so that my skills become decrease"* (Q5), and *"I become lazy to think because I rely on AI"* (Q6) — framing that echoes Al-khresheh's (2024) concern with AI-linked skill degradation, though the present data reflect students' self-perception of this effect rather than an independently measured skill decline.

Output mismatch: Register and coherence

A third, more narrowly linguistic problem was output that students judged unusable as delivered: six participants in the speech task and ten in the interview task described AI-generated language as incoherent or mistranslated, requiring manual correction. One participant described translating AI output only to find *"words were very difficult to understand or pronounce"* (Q3) when performing; another described needing to move AI-translated interview responses into a word processor to manually locate more natural synonyms because *"AI translated them in a language that was sometimes difficult to read"* (Q4). Read as a register-calibration problem rather than a generic technical glitch, this suggests AI tools were producing lexically sophisticated but pragmatically mis-calibrated output (language too formal, too literal, or too unnatural for spoken delivery) that students then had to manually simplify or naturalize before it could function as spoken, not merely written, English. This is a meaningfully different failure mode from the comprehension paradox above: there, students didn't fully process content they could nonetheless reproduce; here, the content itself was not yet fit for the spoken register the task demanded, regardless of whether students understood it.

An absent concern

One further pattern is notable for its absence rather than its presence: no student in this study raised concerns about data privacy or ethical risk in AI use, a marked contrast with Chea and Xiao's (2024) participants. Whether this reflects genuine unconcern, limited awareness of how these tools handle input data, or simply a task framing (speaking performance, not data

handling) that didn't prompt reflection on it, the data here cannot distinguish, but the gap itself is worth noting as a direction for the instrument in future iterations of this research, since its absence was total rather than marginal.

Reading the two majors together

Across these four patterns, BTP and AKT participants differ mainly in degree rather than kind. BTP participants showed both higher AI reliance (25/28 versus 12/30) and higher self-reported comprehension (79.3% interview). The more consistent explanation, given the evidence above, is that self-reported comprehension in this study tracks confidence in handling AI-generated language rather than independently verified command of it: the same BTP cohort with strong self-reported interview comprehension still included over a quarter of speech-task respondents describing shallow understanding, and AKT's mean performance score fell below their own prior baseline despite comparably high self-reported comprehension. Task structure appears to moderate this pattern more than major does: the interview task's fixed linguistic scaffold reduced overreliance-type complaints but surfaced comprehension as the primary concern instead, while the speech task's greater generative demand produced the reverse pattern. For English speaking instruction specifically, this points to a narrower and more actionable implication than "AI helps or harms learning" in general: task design that supplies structured linguistic input, as the interview blueprint did, appears to reduce dependency-related risk more effectively than open AI access paired with post hoc comprehension checks, a pathway future English task design in vocational settings could test directly rather than treating AI access itself as the variable to manage.

Pedagogical implications

Four implications follow directly from the patterns aforementioned each of which is tied to a specific finding rather than AI use in general. First, because heavy AI reliance coexisted with comprehension gaps, most visibly in BTP's speech task, where 28.6% of respondents described shallow or absent understanding despite high self-reported comprehension elsewhere, the participants should be required to demonstrate independent processing of AI-generated content before performance, for instance through mandatory paraphrase or annotation of key passages rather than direct reproduction.

Second, since most participants in both majors began preparation only two to three days before performance despite a full week available, and this compression tracked closely with AI dependency (25/28 BTP, 12/30 AKT), the driver here appears to be workload scheduling rather than the tool itself. Protecting dedicated preparation time is likely to reduce reliance more directly than restricting AI access would.

Third, the interview task, structured around lecturer-provided blueprints, showed both lower overreliance complaints and a different dominant concern (comprehension) than the

more open-ended speech task. This suggests task design can regulate how students draw on AI: tasks with clear linguistic input and models appear to reduce the need to delegate content generation, even without limiting access to AI tools.

Fourth, since no participant raised data privacy or ethical concerns despite extensive AI use, English language learning system incorporating AI should build in explicit guidance on responsible use such as privacy, source integrity, and appropriate delegation rather than assume the learners will develop this awareness independently.

Taken together, these implications suggest that AI's value in English speaking instruction depends less on the tool itself than on how preparation time, task structure, and instructional guidance are designed around it, a conclusion that follows from the task-design contrast (blueprint-supported interview versus open-ended speech) documented above, rather than from AI use as a general phenomenon.

IV. CONCLUSION

This study's contribution lies less in confirming that AI assists ESP preparation than in showing *where the assistance breaks down*. Two findings carry that weight. First, in relation to the initial research question about AI benefits, AI consistently helped the participants shorten the preparation time and raised their confidence. Second, task structure moderated AI dependency more consistently than major did: the blueprint-scaffolded job interview task constrained reliance on AI across both AKT and BTP groups, which is seen as one of the challenges of using AI, while the more open-ended speech task did not. This suggests that instructional scaffolding, not students' disciplinary background, may be the more actionable lever for managing AI dependency in vocational settings. A further observation that was the complete absence of student-voiced concern about data privacy or ethical use, which is also regarded as another challenge of AI implementation, contrasting with the awareness reported in Chea and Xiao's (2024) study indicates that ethical and data-literacy dimensions of AI use are not yet part of these students' frame of reference and warrant deliberate inclusion in English language learning rather than assumption of their presence.

These findings should be read within clear boundaries. The study drew on two intact classes at a single vocational institution, which limits generalizability beyond similar plantation-related programs; it depended substantially on self-reported comprehension and dependency rather than independently rated performance measures; it captured one point in students' academic progression rather than a trajectory; and because students selected their own AI tools, no single platform's features can be credited or blamed for the outcomes observed. Each of these constraints' points to a specific next step: comparative multi-institution designs to test generalizability, mixed-methods designs pairing self-report with rate-based scoring to test the comprehension-performance gap identified here, longitudinal cohort designs to track whether dependency narrows or widens with experience, and tool-controlled designs to

isolate which platform features drive the benefits and challenges reported. Until such evidence accumulates, these results support treating AI as a supportive rather than primary resource in vocational contexts, where task design and instructional guidance appear to shape engagement with AI at least as much as the tool itself.

DECLARATION

Author contribution: AS initiated, wrote, and composed the manuscript. AS and APAT complete the revision of the manuscript sent by the editor/reviewer. APAT finalizes the manuscript according to the review and template from the journal.
Funding statement: No funding.
Conflict of interest: The authors have no conflict of interest related to the manuscript.
Additional information: No additional information is available for this manuscript.

REFERENCES

Adelita, D., & Daulay, S. H. (2023). Postgraduate students' difficulties and strategies on speaking. *NOTION: Journal of Linguistics, Literature, and Culture*, *5*(1), 1-17. https://doi.org/10.12928/notion.v5i1.5827

Ahimsa, A. F. H., & Khawa, D. (2025). The impact of artificial intelligence on English language teaching: Opportunities and challenges in technology era. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 13*(1), 191–201. https://doi.org/10.24256/ideas.v13i1.6270

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. (2023). Impact of ChatGPT on learning motivation: Teachers and students' voices. *Journal of English Studies in Arabia Felix, 2*(1), 41–49. https://doi.org/10.56540/jesaf.v2i1.51

Al-khresheh, M. H. (2024). The future of artificial intelligence in English language teaching: Pros and cons of ChatGPT implementation through a systematic review. *Language Teaching Research Quarterly, 43,* 54–80. https://doi.org/10.32038/ltrq.2024.43.04

Almehmadi, W. (2024). Exploring the potential of AI techniques in teaching English as a foreign language: A systematic literature review. *Asian Journal of Social Sciences and Management Studies, 11*(2), 22–31. https://doi.org/10.20448/ajssms.v11i2.5576

Al-Raimi, M., Mudhsh, B. A., Muqaibal, M. H., & Al-Yafaei, Y. (2024). To what extent has artificial intelligence impacted EFL teaching and learning? A systematic review. *Jurnal Arbitrer, 11*(3), 399–412. https://doi.org/10.25077/ar.11.3.399-412.2024

Bećirović, S., Dubravac, V., & Brdarević-Čeljo, A. (2022). Cooperative learning as a pathway to strengthening motivation and improving achievement in an EFL classroom. *Sage Open, 12*(1). https://doi.org/10.1177/21582440221078016

Chea, P., & Xiao, Y. (2024). Artificial intelligence in higher education: The power and damage of AI-assisted tools on academic English reading skills. *Journal of General Education and Humanities, 3*(3), 287–306. https://doi.org/10.58421/gehu.v3i3.242

Deci, E. L., & Ryan, R. M. (2016). Optimizing students' motivation in the era of testing and pressure: A self-determination theory perspective. In *Building autonomous learners: Perspectives from research and practice using self-determination theory* (pp. 9-29). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-287-630-0_2

Dja'far, V. H., & Hamidah, F. N. (2024). Improving English pronunciation skills through AI-based speech recognition technology. *Ethical Lingua: Journal of Language Teaching and Literature, 11*(2). https://doi.org/10.30605/25409190.747

Ebneyamini, S., & Sadeghi Moghadam, M. R. (2018). Toward developing a framework for conducting case study research. *International journal of qualitative methods*, *17*(1). https://doi.org/10.1177/160940691881795

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review, 61*(4), 5–14. https://doi.org/10.1177/0008125619864952

Halimah, A. N., Budianto, A., Hoesny, M. U., Cabuquin, J. C., & Ulfa, S. (2025). Improving English speaking skills through AI-based voice chatbots: A systematic review of vocational education studies. *Studies in Learning and Teaching*, *6*(3), 785-797. https://doi.org/10.46627/silet.v7i1.684

Kavgić, A. (2024). Effects of using AI in teaching English: A workplace-English case study. *Годишњак Филозофског факултета у Новом Саду, 49*(2), 19–40. https://doi.org/10.19090/gff.v49i2.2534

Kazu, I. Y., & Kuvvetli, M. (2023). The influence of pronunciation education via artificial intelligence technology on vocabulary acquisition in learning English. *International Journal of Psychology and Educational Studies, 10*(2), 480–493. https://doi.org/10.52380/ijpes.2023.10.2.1044

Marwan, A. (2009). ESP teaching challenges in an Indonesian vocational higher institution. *The English Teacher, 38*(1), 1–10. https://meltajournals.com/index.php/TET/article/view/184

Ningrum, F. A., Mulyati, Y., Sumiyadi, S., Halimah, H., & Ye, Z. (2026). The role of artificial intelligence in contemporary literary practice: Perceptions and implications. *NOTION: Journal of Linguistics, Literature, and Culture*, *8*(1), 21–34. https://doi.org/10.12928/notion.v8i1.15292

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin, 134*(2), 270–300. https://doi.org/10.1037/0033-2909.134.2.270

Purbasari, D. A. (2025). Students' perspective and challenge towards utilization of Talkpal.AI in ESP classroom. *ICEETE Conference Series, 3*(1), 18–30. https://doi.org/10.36728/iceete.v3i1.226

Romadhon, R. (2025). Replika AI Chatbot as a tool for enhancing ESP business vocabulary acquisition: A study on polytechnic students. *SALEE: Study of Applied Linguistics and English Education, 6*(1), 183–201. https://doi.org/10.35961/salee.v6i1.1671

Sabrina, A. (2020). The implementation of communicative language teaching in English practice. *NOTION: Journal of Linguistics, Literature, and Culture, 2*(1), 17-27. https://doi.org/10.12928/notion.v2i1.1750

Sariani, Miladiyenti, F., Rozi, F., Haslina, W., & Marzuki, D. (2022). Incorporating mobile-based artificial intelligence to English pronunciation learning in tertiary-level students: developing autonomous learning. *International Journal of Advanced Science Computing and Engineering, 4*(3), 220-232. https://doi.org/10.62527/ijasce.4.3.92

Sheikh, H., Prins, C., & Schrijvers, E. (2023). Artificial intelligence: Definition and background. In *Mission AI: The new system technology* (pp. 15–41). Springer International Publishing. https://doi.org/10.1007/978-3-031-21448-6_2

Siahaan, A., Ardayati, A., & Ramasari, M. (2022). The importance of instrumental motivation among students in teaching English as a foreign language (TEFL) process. *Education of English as a Foreign Language Journal, 5*(1), 1–12. https://doi.org/10.21776/ub.educafl.2022.005.01.01

Stander, M. (2022). The relationship between language learning strategies, affective factors and language proficiency. *Language Learning in Higher Education, 12*(2), 391–408. https://doi.org/10.1515/cercles-2022-2053

Wang, X., & Yan, K. (2022). College English reading teaching integrating production-oriented approach from the perspective of artificial intelligence. *Computational Intelligence and Neuroscience*, 2022, 3411203. https://doi.org/10.1155/2022/3411203

Wass, R., Harland, T., & Mercer, A. (2011). Scaffolding critical thinking in the zone of proximal development. *Higher Education Research & Development*, *30*(3), 317–328. https://doi.org/10.1080/07294360.2010.489237

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology,* 14, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Xatamova, N., & Ashurov, J. (2024). The future of legal English learning: Integrating AI into ESP education. *SPAST Reports, 1*(7). https://doi.org/10.69848/sreports.v1i7.5081

Yang, S., & Wang, W. (2022). The role of academic resilience, motivational intensity and their relationship in EFL learners' academic achievement. *Frontiers in psychology*, 12, 823537. https://doi.org/10.3389/fpsyg.2021.823537

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International journal of educational technology in higher education*, *16*(1), 39. https://link.springer.com/article/10.1186/s41239-019-0171-0