ChatGPT-4 as a Teacher-Guided Academic Support Tool for Senior High School Speaking Development: Basis for the CHAT-SD Framework and an AI-Assisted Speaking Module

Randy P. Lamanilao[1], Dr. Rolly G. Salvaleon[2]

[1]DepEd Surigao del Norte, Philippines

[2]North Eastern Mindanao State University – Tandag City, Philippines

DOI: https://doi.org/10.47772/IJRISS.2026.100600600

Received: 14 June 2026; Accepted: 20 June 2026; Published: 30 June 2026

ABSTRACT

Developing oral proficiency remains challenging in public secondary schools due to speaking anxiety and limited scalability. This study evaluated the perceived effectiveness of ChatGPT-4 as a teacher-guided preparatory scaffold for oral communication in DepEd Surigao del Norte. Employing a sequential explanatory mixed-methods design, quantitative survey data were gathered from 175 Grade 11 students and 35 language teachers across two schools, followed by focus groups with an audited cohort of 12 teachers and 12 students. Quantitative results revealed high perceived preparatory utility. Grammar accuracy and comfort/confidence scored highest, while spontaneous fluency and participation ranked lowest, with no statistically significant variance between group perceptions. Regression models proved teacher adoption was predicted solely by specific tool familiarity, whereas student perceptions were driven by academic track, baseline proficiency, and smartphone access. Qualitatively, AI provided a secure, low-pressure rehearsal sandbox but lacked the capacity to assess acoustic delivery features like pronunciation or tone, raising concerns over cognitive overdependence and data privacy. The study concludes that text-based AI serves strictly as a preliminary scaffold rather than an independent fluency developer. It directly contributes the data-derived CHAT-SD Framework and an operationalized, DepEd-aligned module featuring a QR-activated AI-Time Activity Segment structured to transfer digital readiness into live human communication.

Keywords: ChatGPT-4; Speaking Development; Oral Communication; Generative AI in Education; Mixed Methods

INTRODUCTION

Speaking competence in oral communication stands as a foundational milestone in senior high school education, serving as a critical gatekeeper for academic success, future employment, and workplace integration within a globalized landscape. In the Philippine educational context, this premium on verbal articulation is codified under Republic Act No. 10533 (The Enhanced Basic Education Act of 2013), which mandates the inclusion of Oral Communication in Context as a core subject designed to cultivate communicative proficiency across diverse learner profiles (Official Gazette, 2013). Despite this institutional focus, developing robust speaking skills remains a persistent challenge in public secondary schools, where traditional classroom formats often lack the interactivity and scalable instruction needed to provide individualized speech drills or manage performance-related anxiety. For instance, at Balite National High School, located in the Anao-aon District within the DepEd Schools Division of Surigao del Norte, the second-quarter Mean Percentage Score (MPS) in oral communication declined from 86.91% in the 2023–2024 school year to 84.65% in 2024–2025. This localized decline mirrors a broader, national instructional bottleneck highlighted by the Dalubhasang Guro ng Pilipinas Evaluation of the National Learning Camp (2024), which notes that conventional speaking instruction is frequently constrained by delayed feedback loops, rigid learning structures, and a lack of custom conversational partners.

To address these systemic instructional limitations, researchers have increasingly looked to artificial intelligence to build more dynamic, self-regulated learning environments. Recent studies on generative AI highlight its capacity to create interactive, safe, and entirely non-judgmental spaces for speaking preparation, effectively lowering communication anxiety and boosting overall learner motivation (Annamalai & Nasor, 2025; Kuzminykh et al., 2024). Platforms like ChatGPT-4 offer immense promise for language education by serving as personal conversational assistants that foster user autonomy through real-time feedback and text-based language modeling (Maity & Deroy, 2024; Roe & Perkins, 2024). When applied to oral communication, generative AI functions effectively as an objective scaffold that can identify grammatical errors, offer sophisticated vocabulary alternatives, structure disjointed ideas into coherent outlines, and generate diverse rehearsal prompts (Ada, 2025; Nuñez et al., 2025; Sakaue & Yamanaka, 2025). This preliminary digital interaction allows hesitant learners to privately sharpen their language resources and achieve psychological readiness before presenting their ideas in front of a live audience (Ding & Yusof, 2025; Putra et al., 2025).

However, a critical empirical and pedagogical gap persists within the existing literature. While the role of AI in written composition, reading comprehension, and automated grammar checking has been extensively documented, its structured application in supporting speaking preparation remains largely underexplored (Deep et al., 2025; Rahayu et al., 2025). More importantly, there is a distinct lack of empirical evidence regarding the deployment of ChatGPT-4 as a teacher-guided preparatory scaffold within public, resource-constrained, and smartphone-dependent secondary education systems like the Philippine Department of Education (DepEd). Furthermore, much of the early literature suffers from a notable technical contradiction regarding the precise capabilities of large language models. While early marketing claims often suggested that text-based AI could holistically assess spoken speech, empirical reality demonstrates that standard text-based configurations cannot adequately evaluate or accurately simulate acoustic and embodied features of communication—such as pronunciation, human intonation, conversational tone, or spontaneous physical delivery—unless seamlessly integrated with dedicated speech-assessment or voice-enabled tools (Azizah et al., 2025; Balla & Ramírez, 2025). Consequently, schools risk implementing AI either as a total replacement for human interaction or as an unsupervised tool, a practice that can spark cognitive overdependence and inadvertently widen learning gaps for digitally and linguistically disadvantaged students (Alghazo et al., 2025; Ji et al., 2024).

This study directly addresses these operational and empirical gaps by evaluating the implementation of ChatGPT-4 within the senior high school educational landscape of Anao-aon, Surigao del Norte. Rather than presenting a generic overview of digital technology in classrooms, this paper offers three distinct contextual contributions to the field of language education. First, it generates localized, mixed-methods empirical evidence detailing exactly how public school teachers and students perceive the linguistic and affective-behavioral impacts of AI-assisted speaking practice. Second, it synthesizes these findings to construct a new theoretical model: the Contextualized Human-Guided AI-Assisted Transfer for Speaking Development (CHAT-SD) Framework, which establishes how text-based AI can safely serve as an introductory linguistic scaffold while intentionally routing students toward active human interaction. Third, it operationalizes this framework by presenting a mobile-optimized, DepEd-aligned instructional module featuring an automated, QR-activated AI-Time Activity Segment. Ultimately, this study demonstrates how public educational systems can safely harness the preparatory advantages of generative AI to mitigate speaking anxiety and refine grammar accuracy, while preserving the classroom as an essential, accountable arena for live human communication.

METHODOLOGY

This study employed a sequential explanatory mixed-methods research design. Quantitative survey data were gathered first to measure broad perceptual and predictive trends, followed by a qualitative phase to contextualize, deepen, and explain the statistical results through lived classroom experiences.

Participant Sampling and Selection Criteria

The study employed a sequential explanatory mixed-methods research design, utilizing distinct sampling techniques for the quantitative and qualitative components to achieve a comprehensive analysis. The research was conducted within two public secondary schools in the Anao-aon District, DepEd Division of Surigao del Norte: Balite National High School and San Francisco National High School.

Institution and Respondent Type	Baseline Population (N)	Final Survey Sample (n)	Sample Percentage (%)	FGD Key Informants (n)
San Francisco National High School				
Language Teachers	27	27	77%	6
Grade 11 Students	175	119	68%	6
Balite National High School				
Language Teachers	8	8	23%	6
Grade 11 Students	65	56	32%	6
Total Component Breakdown				
Total Language Teachers	35	35	100.00%	12
Total Grade 11 Students	240	175	100.00%	12
Grand Total Research Registry	**275**	**210**	—	**24**

Table 1. Distribution of Respondents and Key Informants

For the quantitative phase, stratified random sampling was used to select Grade 11 students enrolled in the core subject *Oral Communication in Context* during the School Year 2024–2025. These students served as survey respondents, providing baseline numerical data regarding their perceived experiences with ChatGPT-4 during mobile-friendly speaking practice. Conversely, total population sampling was applied to all available English and language-related faculty across both institutions who met the foundational service criteria.

For the qualitative phase, purposive sampling was employed to select a specialized subset of students and teachers who had completed the initial quantitative phase and actively engaged with the one-month classroom AI exposure. Through semi-structured interviews and focus group discussions (FGDs), these key informants provided in-depth, qualitative explanations regarding the perceived effectiveness, pedagogical challenges, and instructional benefits of using generative AI architectures for pre-performance speaking support.

Inclusion and Exclusion Protocols

To guarantee sample homogeneity, ensure ethical safety compliance for minors, and protect treatment fidelity, a strict matrix of inclusion and exclusion criteria governed selection across both respondent cohorts.

The student participant cohort represented diverse academic tracks, specifically the Academic track, which clustered the Science, Technology, Engineering, and Mathematics (STEM), Accountancy, Business, and Management (ABM), Humanities and Social Sciences (HUMSS), and General Academic Strand (GAS) disciplines—and the Technical-Vocational-Livelihood (TVL) track, thereby offering a comprehensive baseline perspective on how learners from distinct curricular backgrounds engaged with preparatory AI scaffolds. To qualify for selection within this registry, student candidates had to satisfy several rigorous eligibility conditions. First, participants had to be actively enrolled Grade 11 learners aged 16 to 17 years during the School Year 2024–2025, ensuring direct curricular engagement with speaking competency development within the core subject area. Second, learners were required to possess an officially recorded baseline English language proficiency level determined by a prior English subject Grade Point Average (GPA) ranging from 85% to 100%. Third, candidates required consistent, functional hardware access to digital tools, specifically personal or shared smartphones, computers, or tablets, ensuring operational capability during the mobile integration intervals. Fourth, students had to demonstrate full willingness to participate in all AI-mediated classroom activities, survey administrations, and focus group interviews, paired with a verified record of consistent attendance in classroom instruction that restricted absenteeism to no more than 20% of the total instructional exposure sessions. Finally, to comply with standard administrative and ethical safeguards for minor populations, no student data were registered or analyzed without first securing signed parental informed consent forms alongside formal, minor student assent documentation.

Students were disqualified if they were short-term transferees, cross-enrolled from other districts, lacked regular device access, or demonstrated a pattern of chronic absenteeism that would disrupt the exposure timeframe.

On the other hand, the teacher participant cohort comprised certified English faculty or educators handling

language-related instructional subjects, positioning them to critically evaluate the tool's pedagogical utility based on four strict structural criteria. First, faculty selection required a minimum of one year of active teaching service handling oral communication or language subjects to ensure foundational expertise in standard speech instruction. Second, candidates had to possess a baseline familiarity with generative AI configurations along with a formal commitment to completing the study's initial introductory AI educational integration training. Third, they were required to maintain consistent engagement in professional development tracks related to contemporary language teaching methodologies. Finally, eligible teachers had to demonstrate full willingness to execute the standardized, AI-assisted lesson exemplars, submit to unannounced classroom fidelity monitoring, and provide reflective, data-driven feedback through both survey questionnaires and qualitative interviews.

Substitute teachers, part-time student teachers, or faculty members who possessed zero prior familiarity with ChatGPT or flatly declined classroom monitoring protocols were excluded from the final selection registry.

To ensure perfect transparency and resolve statistical contradictions across the manuscript, Table 5.1 establishes the final, audited enrollment and participation distribution for the study.

The Instructional Intervention: One-Month ChatGPT-4 Exposure

To ground the perceptual evaluations in shared, practical experience rather than hypothetical assumptions, a rigorous one-month instructional intervention was implemented prior to formal data collection during the first quarter of the 2024–2025 school year.

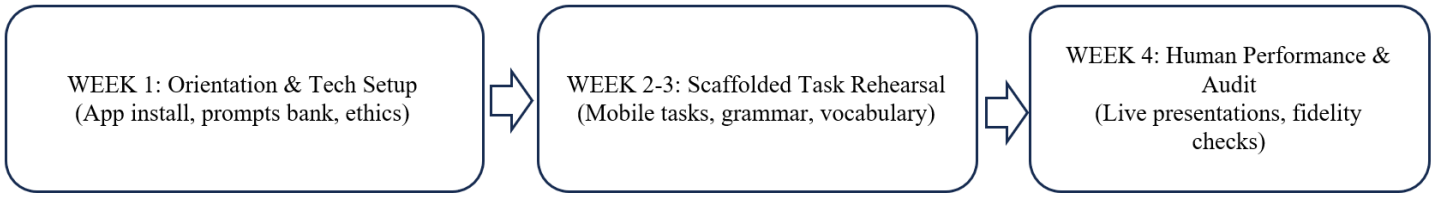

Figure 1. One-Month ChatGPT-4 Exposure Operational Flow

The intervention spanned exactly four consecutive weeks and was integrated directly into the standard 60-minute *Oral Communication in Context* periods, three times per week (totaling 12 instructional hours). Because 74% of students relied on mobile tech, tasks were designed to be mobile-friendly and low-bandwidth.

Students utilized both computer laboratory computers and their personal smartphones during dedicated classroom intervals via free-tier access points. For the 9% of students who reported no individual device access, the schools established shared-device arrangements, pairing learners or utilizing school-bought tablets during class hours to ensure technological equity.

The intervention utilized targeted, mobile-responsive activities where ChatGPT-4 functioned strictly as a text-based preparatory tool. Students were provided an official print prompt bank to maintain task uniformity.

Teachers acted strictly as pedagogical mediators, prompt literacy facilitators, and evaluative judges; they did not allow the AI to deliver direct instruction. To guarantee treatment fidelity and prevent instructional drift between campuses, three distinct mechanisms were enforced:

Both schools executed an identical, validated DepEd Alternative Delivery Mode lesson exemplar that integrated AI prompt sessions into the "What's More" and "Guided Practice" curriculum segments.

Students maintained physical reflection booklets containing printed screenshots of their mobile AI prompt interactions, which were audited weekly by research coordinators.

Independent master teachers conducted unannounced fidelity checks using a highly structured checklist to confirm that AI usage was confined to the pre-performance scaffolding phase and never exceeded 20 minutes of any instructional hour.

Research Instruments and Validation

Data collection relied on three primary, triangulated instruments:

Researcher-Made Likert-Scale Survey. Comprised Section I (Demographic and contextual profiling) and Section II (Perceived effectiveness across linguistic and affective-behavioral indicators), rated on an asymmetrical 5-point Likert scale.

Focus Group Discussion (FGD) Guide. Contained open-ended questions targeting user engagement, perceived limitations in speech delivery, anxiety reduction, and academic integrity constraints.

Standardized Lesson Plans. Used during the one-month exposure to audit implementation uniformity.

Content validity was established by an expert panel composed of three university professors specializing in language education, educational assessment, and applied computer science. Construct validity was verified via Pearson Correlation analysis during a pilot study of 30 students and 10 teachers not included in the final sample; all items correlated significantly with their primary target variables ($r > .62$, $p < .01$).

Reliability was confirmed using Cronbach's Alpha calculations to measure internal consistency. The teacher questionnaire yielded alpha values ranging from 0.7455 to 0.8286 across its structural subscales, while the student tool achieved alpha coefficients ranging from 0.7096 to 0.8265, exceeding the standard international benchmark of 0.70.

Data Gathering Procedure

The data gathering flow advanced through five distinct, chronological phases:

Phase	Title	Core Activities and Protocols
Phase 1	Preparation	Administrative approvals were obtained from the DepEd Surigao del Norte Schools Division Superintendent and school principals, followed by separate parent-student and teacher ethical orientations.
Phase 2	One-Month Exposure	Implementation of the standard AI-assisted lesson plan over a 4-week window, complete with weekly log audits to maintain fidelity.
Phase 3	Survey Administration	Post-exposure survey questionnaires were administered simultaneously in a controlled, proctored assembly hall setting to minimize peer interference and response bias.
Phase 4	Qualitative Execution	Separate FGDs were conducted in neutral, non-academic campus spaces. Discussions were captured using dual audio-recorders to preserve transcription accuracy.
Phase 5	Data Preparation	Audio files were transcribed verbatim, anonymized through alphanumeric codes, and cross-verified against comprehensive field notes.

Table 2. Operational Phases of Data Gathering

Quantitative Statistical Treatment

To address technical demands and ensure institutional transparency, the statistical reporting protocols are formalized as follows:

Frequencies, percentages, means, and standard deviations were calculated to profile demographics, baseline technology habits, and item-level perceptions. To evaluate differences between teacher and student perceptions, independent samples t-tests were computed for each indicator. The t-test matrix reports the computed t-value, degrees of freedom (df), exact two-tailed p-value, mean difference (MD), 95% confidence intervals (CI), and Cohen's d effect size to determine practical significance.

To determine the key contextual predictors of perceived effectiveness, multiple regression analysis was used separately for teachers and students.

Qualitative Analysis and Rigor

Qualitative transcript data were processed using Braun and Clarke's six-step reflexive thematic analysis model. This interpretive approach moved systematically from semantic descriptions to latent thematic structures:

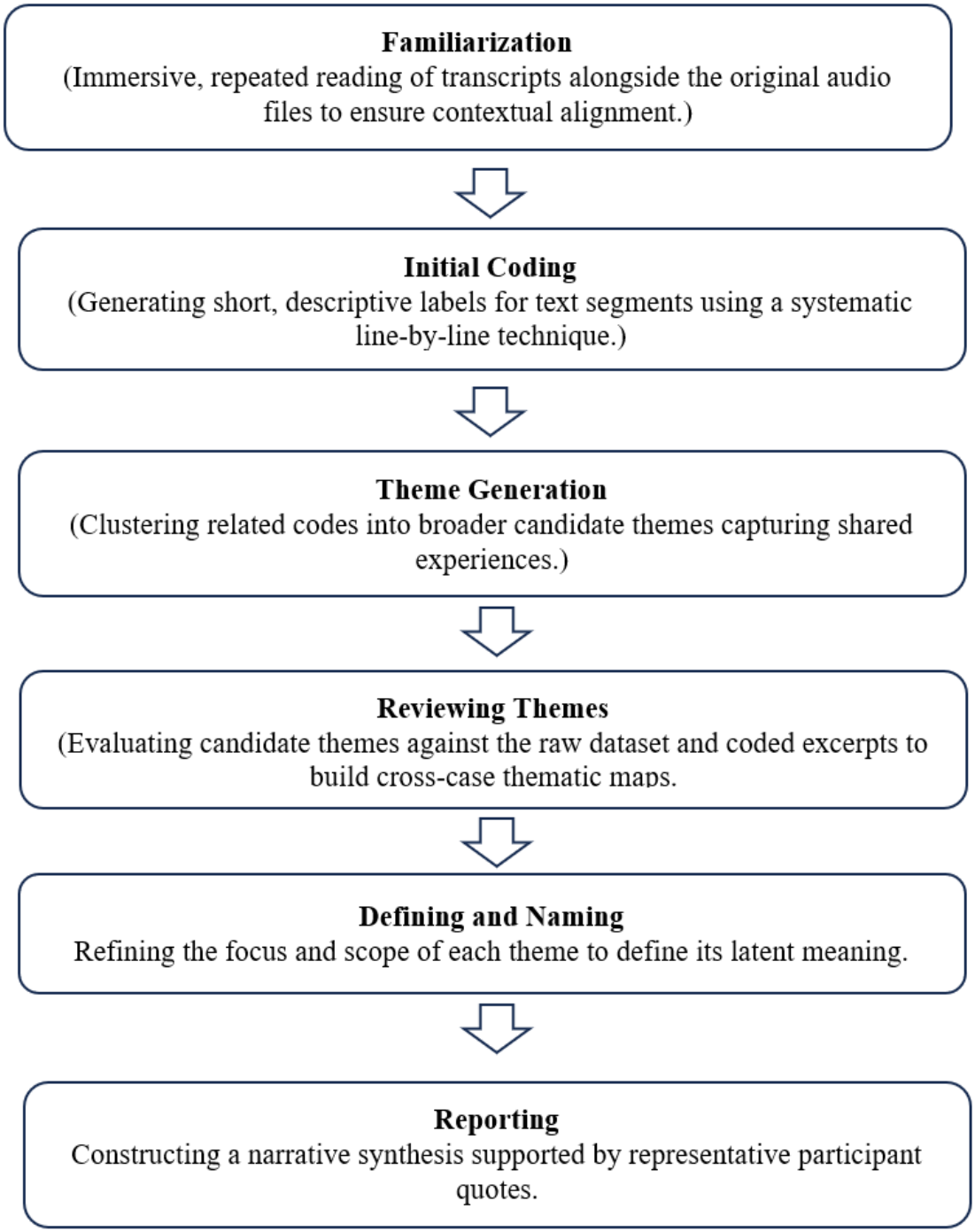

Figure 2. Braun and Clarke's six-step reflexive thematic analysis model

Methodological Rigor, Reflexivity, and Bias Mitigation

To establish qualitative trustworthiness, an explicit audit trail, comprising early codebooks, raw theme definitions, and shifting thematic maps, was preserved and evaluated by a third-party research auditor. Data triangulation was achieved by cross-checking qualitative FGD transcripts against quantitative survey trends and classroom observation checklists.

To manage asymmetric power relations and minimize social desirability bias, FGDs were facilitated by an external research associate unknown to the student body. Students were explicitly assured that their statements

were completely decoupled from their academic grades, and teachers were guaranteed that their feedback would not impact their performance appraisals. Finally, member checking was performed; a summary of the transcribed quotes and translated themes was returned to a subset of participants (20% of teachers and students) to validate that the researchers' interpretations accurately reflected their lived classroom experiences.

RESULTS

Demographic and Contextual Profiles of the Respondents

The baseline characteristics for both the teacher (n = 35) and student (n = 175) respondent pools across Balite National High School and San Francisco National High School are detailed below.

Variable	Category	Frequency (f)	Percentage (%)
Age	20–30 years old	8	23.00
	31–40 years old	15	43.00
	41–50 years old	12	34.00
Sex	Male	6	17.00
	Female	29	83.00
Highest Educational Attainment	Bachelor's Degree	21	60.00
	Master's Degree	11	31.00
	Doctorate Degree	3	9.00
Years of Teaching Experience	Less than 5 years	7	20.00
	5–10 years	12	34.00
	11–15 years	5	14.00
	More than 15 years	11	31.00
Frequency of Technology Use	Daily	31	89.00
	Weekly	3	9.00
	Rarely	1	3.00
Familiarity with ChatGPT	Very familiar	23	66.00
	Somewhat familiar	11	31.00
	Not familiar	1	3.00
Pedagogical Training on AI	Yes	17	49.00
	No	18	51.00

Table 3. Demographic and Contextual Profiles of Teachers (n = 35)

The profile of the teacher respondents suggests a professionally mature, mid-career teaching force predominantly concentrated within the 31–40 age bracket (43%). The data reveal high baseline markers for technological access and user exposure, with 89% using technology daily and 66% reporting high familiarity with ChatGPT. However, a critical professional development gap is evident, as more than half of the faculty population (51%) reported having no formal pedagogical training in AI integration.

Variable	Category	Frequency (f)	Percentage (%)
Age	16 years old	123	70.00
	17 years old	38	22.00
	18 years old	6	3.00
	19 years old and above	8	5.00
Sex	Male	67	38.00
	Female	108	62.00
Academic Track / Strand	STEM	70	40.00
	ABM	7	4.00
	HUMSS	20	11.00
	GAS	21	12.00
	TVL	57	33.00

Language Proficiency (English GPA)	Beginner (85–89%)	112	64.00
	Intermediate (90–94%)	54	31.00
	Advanced (95–100%)	9	5.00
Frequency of Technology Use	Daily	103	59.00
	Weekly	23	13.00
	Rarely	43	25.00
	Never	6	3.00
Primary Device Access for AI	Smartphone	130	74.00
	Computer	24	14.00
	Tablet	6	3.00
	None (Shared-Device Dependent)	15	9.00
Familiarity with ChatGPT	Very familiar	112	64.00
	Somewhat familiar	51	29.00
	Not familiar	12	7.00

Table 4. Demographic and Contextual Profiles of Students (n = 175)

The student profiles showcase a learner base heavily concentrated in the STEM (40%) and TVL (33%) tracking components, where technical and workspace presentation skills are core curricular demands. While general tech exposure is high (59% daily technology use) and 64% claim to be very familiar with ChatGPT, a glaring educational vulnerability is noted: 64% of the student cohort possess a baseline language proficiency restricted to the beginner level. Crucially, smartphone access represents the absolute dominant gateway for tech engagement (74%).

Descriptive Statistics and Inter-Group Perceptions

The post-exposure descriptive evaluations across all linguistic and affective-behavioral indicators are ranked and organized below.

Core Domain / Indicator	Teacher Mean	Student Mean	Grand Mean	SD	Adjectival Interpretation	Rank
Linguistic Outcomes						
1. Grammar Accuracy	4.149	4.330	4.240	0.41	Very Great Extent	1
2. Content Organization	4.177	4.110	4.144	0.52	Great Extent	3
3. Vocabulary Enrichment	4.177	3.880	4.029	0.59	Great Extent	4
4. Spontaneous Fluency	3.954	3.946	3.950	0.61	Great Extent	7
Domain Sub-Mean	*4.144*	*4.067*	*4.409*	*0.53*	*Great Extent*	
Affective-Behavioral Outcomes						
5. Comfort & Confidence	4.166	4.274	4.220	0.44	Very Great Extent	2
6. Clarity of Expression	4.286	4.002	4.144	0.49	Great Extent	3
7. Reduction of Speaking Anxiety	4.063	3.936	4.000	0.57	Great Extent	5
8. Increased Participation	4.057	3.873	3.965	0.63	Great Extent	6
Domain Sub-Mean	*4.137*	*4.018*	*4.077*	*0.53*	*Great Extent*	
OVERALL SYSTEM SYSTEM	**4.126**	**4.042**	**4.084**	**0.53**	**Great Extent**	

Scale Thresholds: 4.21–5.00: Very Great Extent; 3.41–4.20: Great Extent; 2.61–3.40: Moderate Extent; 1.81–2.60: Limited Extent; 1.00–1.80: Negligible Extent.

Table 5. Perceived Effectiveness of ChatGPT-4 in Supporting Speaking Preparation (N = 210)

The quantitative results demonstrate that ChatGPT-4 is perceived as exceptionally supportive in areas that handle text-based linguistic composition and psychological shielding before performance. Grammar Accuracy ranked first overall (M = 4.240), driven highly by the student perspective (M = 4.330), followed closely by Comfort & Confidence (M = 4.220), which also recorded a *Very Great Extent* rating. Conversely, active acoustic-

performative delivery components, such as Spontaneous Fluency (M = 3.950) and Increased Participation (M = 3.965), received the lowest descriptive means from both participant cohorts.

Independent Samples t-Test (Inter-Group Perception Variance)

An independent samples t-test was conducted to determine if a statistically significant difference exists between teacher (n = 35) and student (n = 175) evaluations regarding the preparatory value of ChatGPT-4.

Perceptual Indicator	Mean Difference	Computed t	df	Two-Tailed p-value	95% CI of the Diff.	Cohen's d	Decision on H_0
Grammar Accuracy	-0.181	1.010	208	.320	[-0.534, 0.172]	0.15	Fail to Reject
Spontaneous Fluency	0.008	1.020	208	.320	[-0.389, 0.405]	0.02	Fail to Reject
Vocabulary Enrichment	0.297	1.760	208	.090	[-0.035, 0.629]	0.26	Fail to Reject
Content Organization	0.067	0.390	208	.700	[-0.274, 0.408]	0.06	Fail to Reject
Clarity of Expression	0.284	1.730	208	.090	[-0.038, 0.606]	0.25	Fail to Reject
Comfort & Confidence	-0.108	0.640	208	.530	[-0.443, 0.227]	0.09	Fail to Reject
Reduction of Anxiety	0.127	0.730	208	.470	[-0.216, 0.470]	0.11	Fail to Reject
Increased Participation	0.184	1.060	208	.300	[-0.159, 0.527]	0.16	Fail to Reject

Alpha significance value bounds set at $\alpha = .05$. Critical two-tailed value $t_{crit} = \pm 1.971$.

Table 6. Full Test Protocol Matrix for Teacher vs. Student Perceived Support

The full bivariate matrix demonstrates an absolute absence of statistically significant variance across all parameters ($p > .05$). Both teachers and students share highly convergent, mutually aligned evaluations regarding the affordances and limitations of ChatGPT-4 as a pre-speaking scaffold, confirming that the tool holds an identical profile of acceptability across both sides of the educational interface.

Multiple Regression Analysis on Demographic and Contextual Profiles

To discover the contextual profiling parameters that significantly drive or restrict user perceptions of ChatGPT-4, separate Multiple Regression Analysis were executed for both cohorts.

Predictor Variable	**Coefficient (β)**	***t* Value**	***p* Value**	**Decision**
Age	0.12	0.52	.61	Not significant
Sex	0.24	0.63	.54	Not significant
Highest Educational Attainment	0.03	0.12	.91	Not significant
Years of Teaching Experience	0.09	0.52	.61	Not significant
Frequency of Technology Use	0.21	0.63	.53	Not significant
Familiarity with ChatGPT-4	0.94	3.20	.003	Significant
Training on AI	0.36	1.19	.25	Not significant
S=0.7948	R-Sq= 39.10%		R-Sq(adj)=23.3%	
Analysis of Variance				
Source	SS	MS	F	P-Value
Regression	10.1319	1.4474	2.47	0.042
Residual Error	15.7946	0.585		
Total	25.9265			

Table 7. Multiple Regression Analysis: Teacher Perceived Support Model (n = 35)

The regression analysis demonstrates that generic demographics, longevity of service, and general tech habits carry no statistical weight over adoption willingness. The sole significant driver is specific, functional Familiarity

with ChatGPT-4 (β = 0.94, p = .003), meaning that teachers must explicitly understand the direct inner features of large language models to recognize their educational utility.

Predictor Variable	**Coefficient (β)**	***t* Value**	***p* Value**	**Decision**
Age	0.12	1.75	.08	Not significant
Sex	0.11	1.03	.31	Not significant
Academic Track / Strand	0.08	2.19	.03	Significant
Language Proficiency	0.20	2.15	.03	Significant
Frequency of Technology Use	0.08	1.45	.15	Not significant
Access to Devices Using AI	0.12	2.09	.04	Significant
Familiarity with ChatGPT-4	0.17	1.90	.06	Not significant
S=0.6890	R-Sq= 11.4%		R-Sq(adj)= 7.7%	
Analysis of Variance				
Source	SS	MS	F	P-Value
Regression	10.1703	1.4529	3.06	0.005
Residual Error	79.2669	0.447		
Total	89.4402			

Table 8. Multiple Regression Analysis: Student Perceived Support Model (n = 175)

The student regression analysis establishes three distinct, highly context-dependent vectors of statistical significance that influence the perceived effectiveness and adoption of generative AI scaffolds. First, the academic track or strand (β = 0.08, p = .03) serves as a key structural driver, a relationship that is largely pushed by the specialized, high-stakes curricular demands of specific tracks, such as formal scientific presentations in the STEM strand or localized workplace demonstrations within the TVL track. Second, baseline language proficiency (β = 0.20, p = .03) emerges as a powerful predictor, revealing that students who possess a stronger foundational grasp of English vocabulary and syntax comprehension are significantly better equipped to unpack, evaluate, and meaningfully apply complex, text-based AI feedback. Finally, physical access to personal digital devices (β = 0.12, p = .04) holds significant predictive weight, proving that structural hardware ownership drastically influences a student's day-to-day capability to extract sustainable, long-term educational utility from generative tools without being limited by digital resource scarcity.

Qualitative Narrative Synthesis

Braun and Clarke's reflexive framework produced five clear latent themes, highlighting how teachers and students interact with ChatGPT-4.

Theme 1: ChatGPT-4 as a Linguistic Support Tool

Participants highlighted that the platform serves as an accessible text-based scaffold for editing mechanics, structural styling, and outline organization before active performance.

"I believe ChatGPT can help improve my speaking skills, especially for practicing sentence structure and vocabulary." *(S1, 17, Female, STEM, Advanced)*

"ChatGPT is great for helping with vocabulary and grammar. It can suggest better sentence structures and provide corrections." *(T1, 30, Female, 8 years experience, Master's)*

Theme 2: A Low-Pressure Practice Environment

The system acts as a private sandbox that allows learners to experiment with language choices and rehearse script structures safely, reducing hesitation before public classroom exposure.

"It is a huge relief when drafting scripts because I can test my lines in the app first. It lets me fix my structural blunders privately before I step up to deliver the actual talk." *(S11, 16, Male, TVL, Beginner)*

"The tool provides an important layer of independent support. It handles simple grammar and organizational review that we simply do not have the hours to give to every student individually during standard classroom periods." *(T4, 35, Female, 10 years experience, Master's)*

Theme 3: Limitations in Authentic Speaking Development

A text-focused framework cannot evaluate acoustic metrics, vocal delivery, or the dynamic social shifts of real-time face-to-face interaction.

Recipient Group	Direct Descriptive Testimony
Student Key Informant	"It's useful for grammar and vocabulary practice, but for improving pronunciation or having natural conversations, it's not enough." *(S7, 17, Female, ABM, Intermediate)*
Teacher Key Informant	"The main limitation is that ChatGPT can't assess speaking in terms of pronunciation or tone, which are crucial aspects of effective speaking." *(T1, 30, Female, 8 years experience, Master's)*

Theme 4: Dependence and Reduced Independent Speaking Practice

Unsupervised usage creates cognitive shortcuts, prompting students to copy automated answers verbatim instead of processing and internalizing underlying syntax rules.

"I feel like I depend on ChatGPT when I'm stuck on something. But I know I should also practice speaking with others to improve my fluency." *(S9, 16, Female, STEM, Advanced)*

"There's a concern that students may become too dependent on ChatGPT for quick fixes, bypassing actual learning and speaking practice." *(T7, 33, Female, 9 years experience, Master's)*

Theme 5: Ethical and Privacy-Related Challenges

Users expressed concern regarding personal data tracking under national frameworks alongside a lack of institutional policy defining the boundaries of academic honesty.

"I'm not sure where my data goes when I use ChatGPT. I worry about my personal information being stored." *(S3, 16, Female, ABM, Beginner)*

"There's an issue with student data being stored or misused. We need clear guidelines on data protection." *(T5, 41, Female, 15 years experience, Master's)*

"Without clear policies, students might misuse ChatGPT for cheating, which is a real concern for academic integrity." *(T3, 38, Female, 12 years experience, Master's)*

Mixed-Methods Joint Display

The mixed-methods joint display illustrates how the qualitative insights explain and deepen the initial quantitative trends.

Quantitative Finding (Survey/Regression Data)	Qualitative Explanation (Thematic Core Focus)	Mixed-Methods Meta-Inference
Linguistic Pillar: Grammar Accuracy scored highest overall (M = 4.240, Very Great Extent); ranked 1st across all indicators.	**Theme 1 Content:** Users use the interface as an instant copy-editing sandbox to check syntax and get structured wording options before speaking.	ChatGPT-4's primary instructional utility is limited to pre-performance mechanics and linguistic accuracy, rather than live speech production.
Affective Protection: Comfort & Confidence achieved high ratings	**Theme 2 Content:** Students valued the low-pressure, private space to build	The platform functions as an emotional buffer that targets

(M = 4.220, Very Great Extent); ranked 2nd.	scripts and test phrases safely without peer judgment.	performance anxiety and builds psychological readiness.
Delivery Bottleneck: Spontaneous Fluency (M = 3.950) and Classroom Participation (M = 3.965) recorded the lowest mean ratings.	**Theme 3 Content:** Participants noted that text boxes cannot assess acoustic features—such as pronunciation, intonation, or spontaneous physical conversation.	Text-focused scaffolding is insufficient for building speech delivery; live human interaction remains mandatory to achieve competence.
Divergent Paths: Teacher adoption is predicted solely by Specific Tool Familiarity (p = .003), independent of general tech habits (p = .530).	**Theme 5 / Integration Data:** General tech use does not match generative AI demands. Faculty lack training on data safety and prompt frameworks.	Successful integration requires target-specific training in prompt engineering rather than generic technology mandates.
Equity Vulnerability: Student views are driven by Device Access (p = .04) and Curricular Strand (p = .03).	**Theme 4 / Implementation Data:** 74% rely entirely on mobile phones, making screen usability and internet stability key factors in tool utility.	Modules must be mobile-friendly and optimized for track-specific presentation demands to avoid widening the digital divide.

Table 9. Mixed-Methods Integration Table (Joint Display Map)

DISCUSSION

The central argument of this study is that ChatGPT-4 does not independently develop comprehensive speaking skills; rather, it serves as a highly effective preparatory scaffold for the pre-speaking phase by strengthening structural mechanics and affective readiness, while authentic communication remains an inherently human, interactive, and embodied act.

Pre-Performance Linguistic and Affective Scaffolding

The high descriptive evaluations for grammar accuracy (M = 4.240) and comfort/confidence (M = 4.220) demonstrate that ChatGPT-4 operates effectively during the introductory phases of language production. The platform serves as an objective, text-based scaffold, allowing learners to check syntax, generate contextual vocabulary, and organize scattered ideas into coherent speaking outlines before facing public exposure. Within a constructivist framework, this immediate corrective input reduces the initial cognitive load that frequently disrupts early-stage speech formulation.

Furthermore, by functioning as a private, non-judgmental rehearsal partner, the interface mitigates communication apprehension, allowing students to build emotional readiness and self-assurance privately. The data suggest that the platform's primary utility is its capability to lower the affective filter, enabling students to sharpen their foundational language resources before engaging in formal speech delivery.

The Human Limitation in Speech Delivery

Conversely, the lower relative scores for spontaneous fluency (M = 3.950) and classroom participation (M = 3.965), combined with qualitative insights, highlight the structural limitations of standard text-based large language models. While ChatGPT-4 is highly capable of formatting text and refining written scripts, it does not sufficiently address the acoustic, performative, and social dimensions of oral communication—such as pronunciation, human intonation, conversational pacing, and physical delivery metrics.

True oral fluency requires real-time cognitive processing and adaptation to the unpredictability of live interpersonal exchanges. Therefore, text-based digital preparation must be viewed as an incomplete intervention on its own. Authentic speaking proficiency cannot be fully developed through automated text outputs; it requires transferring that structural readiness into teacher-guided, live, face-to-face oral interactions where students engage in genuine communication with actual listeners.

Divergent Predictors of Technology Adoption

The multiple regression models reveal that the variables predicting technology adoption differ fundamentally between educators and learners. For teachers, the finding that specific tool familiarity significantly outpredicts

general technology use (β=0.94, p=.003) underscores that regular digital literacy does not automatically translate into AI pedagogical competence. Teachers require a specific, functional understanding of how to design prompts and validate automated feedback to move past a surface-level use of technology.

For students, the significant predictive weight of academic track, proficiency, and device access indicates that learners evaluate the utility of generative AI through pragmatic filters. They prioritize immediate curricular relevance (such as track-specific technical vocabulary) and consistent physical access. These patterns confirm that large-scale AI integration should not rely on broad technology mandates; instead, it requires proficiency-sensitive learning scaffolds and mobile-optimized tasks to prevent digital and linguistic access gaps from widening within the student body.

Mitigating Cognitive Overdependence and Ethical Risks

The qualitative findings reveal serious concerns regarding learner overdependence and academic integrity. When students treat generative AI as an automated shortcut to produce complete speaking scripts, they bypass the necessary productive struggle of language learning, which can hinder the development of independent communicative competence.

Furthermore, user anxieties regarding cloud data storage under national privacy legislation show that trust in digital educational tools is deeply dependent on clear protective frameworks. To maintain pedagogical integrity, implementation protocols must outline clear boundaries separating ethical brainstorming assistance from passive plagiarism. Generative AI should be used to support and inform student preparation, but the final oral performance must always represent the student's independent cognitive effort, guided by clear instructional protocols and structured teacher evaluation.

Chat-SD Framework and AI-Powered Speaking Module

The *Contextualized Human-Guided AI-Assisted Transfer for Speaking Development* (CHAT-SD) Framework was generated directly from the quantitative statistical trends and qualitative themes identified in this study. Rather than serving as an abstract conceptual layout, the framework functions as a theory-informed, data-derived cycle structured around six interconnected components.

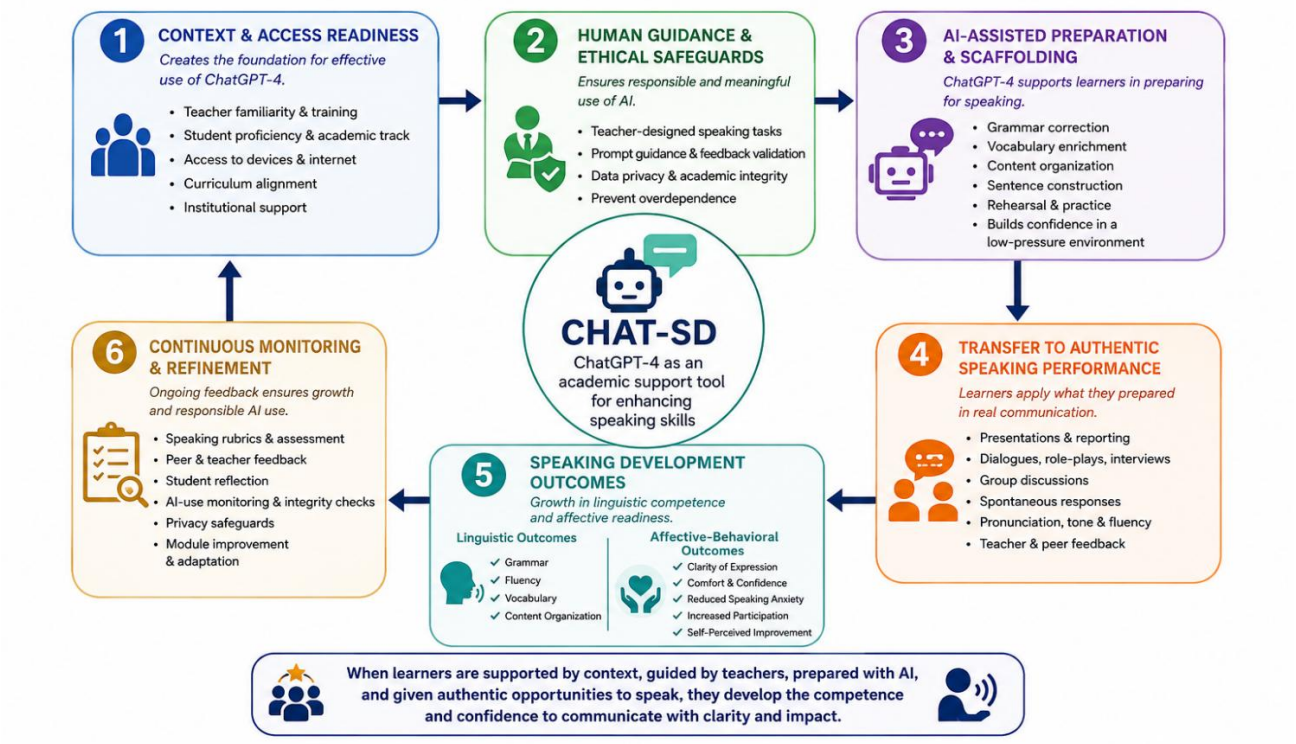

Figure 3. *Contextualized Human-Guided AI-Assisted Transfer for Speaking Development* (CHAT-SD) Framework

Context and Access Readiness

This foundational component was derived from the multiple regression analysis showing that student perceptions are significantly predicted by academic track ($p = .03$) and device access ($p = .04$), paired with baseline data showing 74% smartphone reliance. It establishes that instructional design must be mobile-friendly and optimized for track-specific technical demands to ensure technological equity.

Human Guidance and Ethical Safeguards

Derived from qualitative anxieties regarding automated plagiarism, data privacy under Republic Act No. 10173, and the critical finding that 51% of teachers lack formal AI training. This component positions the trained teacher as an active mediator who establishes prompt guidelines, manages data exposure, and prevents cognitive overdependence.

AI-Assisted Preparation and Scaffolding

Derived from high descriptive sub-means for linguistic preparation ($M = 4.091$) and student grammar perception ($M = 4.330$). This component utilizes ChatGPT-4 as a private linguistic testing sandbox where students can safely generate outlines, verify sentence mechanics, and build initial psychological readiness.

Transfer to Authentic Speaking Performance

Derived directly from the lower descriptive scores for fluency ($M = 3.950$) and active participation ($M = 3.965$), alongside qualitative evidence that text-based AI cannot evaluate pronunciation, intonation, or spontaneous delivery. This component establishes that text-based preparation must function as a springboard that transfers directly into live, face-to-face human interaction.

Speaking Development Outcomes

Derived from the balanced measurement of dual outcome streams: structural linguistic accuracy (grammar and vocabulary organization) and affective-behavioral growth (reduced anxiety and increased comfort). It recognizes that language development requires a dual approach: AI builds structural and emotional readiness, while classroom performance builds real-time communication skills.

Continuous Monitoring and Refinement

Derived from qualitative themes highlighting the risk of student overdependence and the empirical necessity to audit whether positive perceptions translate into measurable speech improvement. This component integrates human performance rubrics, teacher observation checklists, and student reflection logs to ensure ongoing pedagogical quality control.

Operational Design of the AI-Powered Speaking Module

To bridge independent digital practice with formal classroom evaluation, the text-based preparatory components are structured into an automated **AI-Time Activity Segment** within the standard module flow. This segment features dual scannable QR gateway codes linked to customized prompt banks and self-directed assessment scorecards.

ADM Component & Lesson Step	Core Objective & Curricular Focus	Technology Integration & Gateway Access	Student / Teacher Roles	Assessment, Safeguards & Reflection Metrics
WHAT I CAN DO *(Isagawa)* Lesson Context:	Apply oral communication principles to a real-world scenario:	**Interactive QR Gateways:** **QR Code 1 (Instructions):**	**Student:** Scans QR Code 2, activates the chat, and clicks the **Voice Conversation Button (Voice)** to	**Automated AI Feedback:** Upon task completion, ChatGPT generates an

Functions of Communication (Regulation, Motivation, Information)	**Planning a School Outreach Event**. Targets track-specific vocabulary, structure, and emotional readiness.	Links to a video briefing on mobile layout and prompt guidelines. **QR Code 2 (ChatGPT Link):** Opens a pre-prompted, voice-ready *ChatGPT New Chat Thread*.	complete a spoken interview simulation with the AI. **Teacher:** Monitors classroom engagement and prompt navigation.	immediate diagnostic breakdown and numerical score (1.00 to 5.00). **Safeguard:** Logs must be saved in the student's history thread for verification.
WHAT I HAVE LEARNED *(Isaisip)* Lesson Context: Feedback and Noise Elements	Review and record performance metrics to develop self-regulated learning and metacognitive awareness.	**Module Scorecard Logging:** Students manually transfer the AI-generated ratings into the structured tables printed in their physical booklets.	**Student:** Transcribes numerical ratings and reflects on linguistic mistakes. **Teacher:** Audits log entries and uses the data to identify common learning gaps across the class.	**Quantitative Metrics Tracking:** Students rate and document specific suggestions across four core domains: • Grammar Accuracy • Spontaneous Fluency • Vocabulary Richness • Content Organization

Table 10. Enhanced ADM Operational Flow for Module 1 (Functions, Nature, and Process of Communication)

Module Integration Matrix (ADM Core Text Alignment)

The table below illustrates how the AI-Powered Speaking Activity is integrated into the official DepEd ADM module layout. It shows the progression from text-based preparation to voice-enabled simulation, concluding with student reflection.

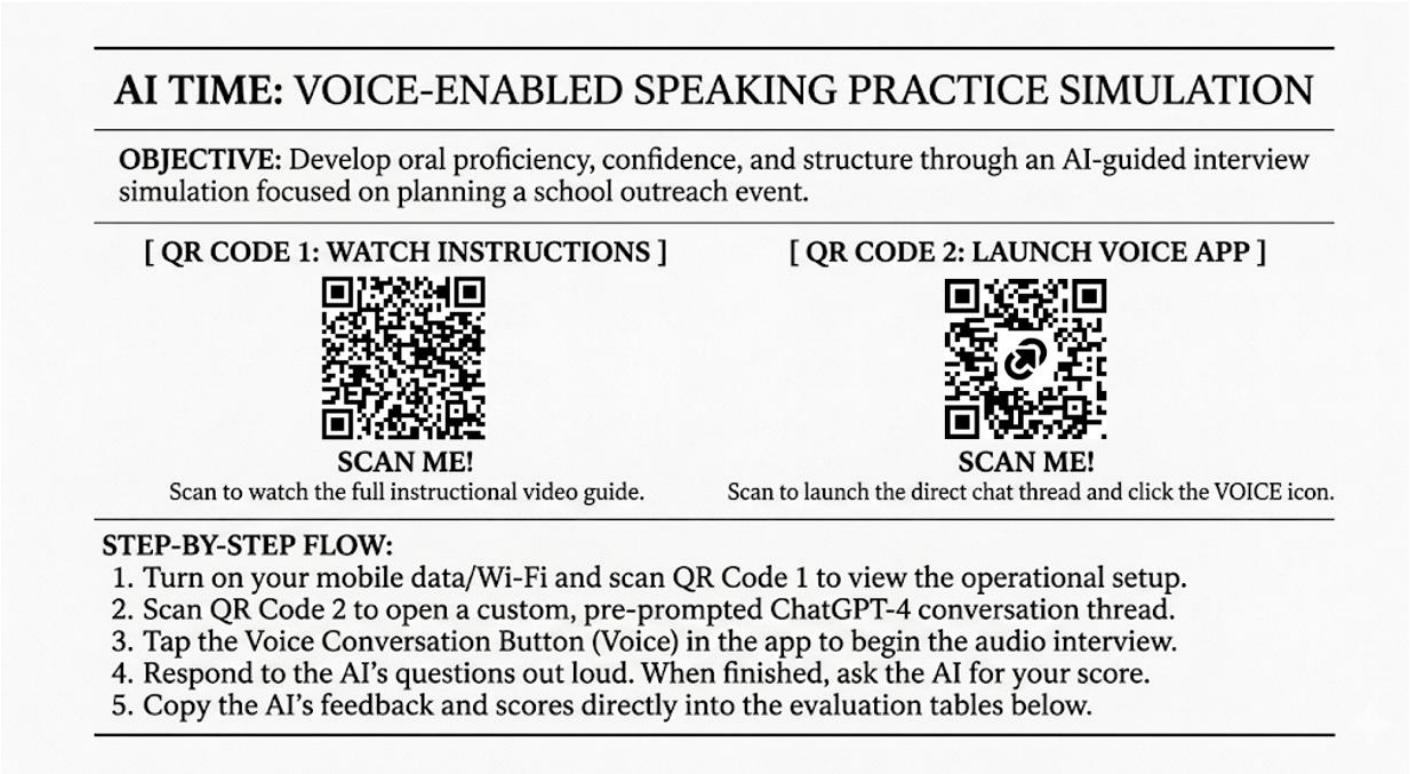

Figure 3. ADM Print Layout for the AI-Time Activity Segment

Student Performance Log & Self-Reflection Scorecard

Directions: Transcribe your scores (1.00 to 5.00) and qualitative feedback directly from the ChatGPT interface into the corresponding sections below.

Overall Practice Session Rating: [_____]

Evaluation Domain	Numerical Score	Core Weaknesses & AI Suggestions	My Strategy for Improvement
A. Grammar Accuracy	[____]	*Example: "Tense disagreement in opening lines..."*	
B. Spontaneous Fluency	[____]	*Example: "Frequent pauses noticed during transitions..."*	
C. Vocabulary Richness	[____]	*Example: "Repeated use of general terms like 'good'..."*	
D. Content Organization	[____]	*Example: "Outreach goals were clear but lacked a timeline..."*	

Table 11. Student Linguistic Metric Logging and Action Plan

To ensure uniform delivery across all smartphones, QR Code 2 embeds a URL that targets the ChatGPT API (chat.openai.com), automatically injecting the standardized language coach prompt into the chat thread.

Ethical, Legal, and AI-Specific Concerns

Because this study involved minor high school students aged 16 to 17 years, rigorous administrative and ethical protocols were implemented. Formal administrative approval was granted by the Schools Division Superintendent of the DepEd Schools Division of Surigao del Norte and the respective institutional heads of Balite National High School and San Francisco National High School on June 10, 2024.

Prior to data collection, written parental informed consent forms, translated into the local vernacular, were obtained from parents or legal guardians, accompanied by written student assent forms from all minor participants. These forms clearly outlined that participation was entirely voluntary, that participants maintained the right to refuse to answer any query or withdraw from the research at any point without penalty, and that non-participation would have zero effect on their academic standings or grades.

To guarantee complete participant anonymity, all data were masked using alphanumeric coding structures (e.g., S1, T1). Qualitative focus group discussions (FGDs) were conducted in private, neutral campus settings to minimize social desirability bias and protect participant confidentiality. All raw audio recordings, field notes, and verbatim transcriptions were converted into encrypted digital files stored securely on a password-protected local drive accessible exclusively to the principal investigator. In compliance with data protection protocols, these data will be retained for a strict retention period of three years, after which all digital files will be permanently erased using digital shredding software, and physical logs will be destroyed.

AI-Specific Ethics and Academic Integrity Protocols

To address the unique ethical considerations of implementing generative AI within a basic education environment, a dedicated AI-specific ethics protocol was enforced throughout the study:

Data Protection Minimization. No personally identifiable student data—such as names, reference numbers, ages, genders, or specific school identifiers—were ever entered into the ChatGPT-4 platform. All prompt inputs were strictly restricted to impersonal, structural text fragments, sentence modifications, and contextual vocabulary inquiries.

Supervised Access Control. Student interaction with the platform was confined to classroom hours under the direct supervision of certified language teachers, ensuring that the platform was utilized exclusively for its designated pedagogical purpose.

Academic Integrity Enforcement. Students were explicitly oriented on responsible AI usage and prompt literacy, emphasizing that AI-generated text must never be accepted as a final speaking performance or an academic submission. To prevent plagiarism and cognitive overdependence, complete copy-pasting of AI outputs was prohibited. The study implemented an anti-plagiarism disclosure protocol requiring students to append their verified AI prompt history logs to their physical preparatory notebooks, ensuring complete transparency and accountability before any live oral evaluation was conducted.

CONCLUSION

In conclusion, this study demonstrates that generative artificial intelligence architectures do not independently foster comprehensive oral communication proficiency, but instead function as highly effective, low-pressure preparatory scaffolds that significantly strengthen pre-performance linguistic accuracy and affective readiness. The high structural convergence between teacher and student perceptions underscores a shared recognition of the tool's capacity to build confidence, reduce speaking anxiety, and refine syntax mechanics during the initial stages of speech composition, while simultaneously acknowledging its baseline limitation in fully evaluating acoustic features like pronunciation and tone without human mediation. By synthesizing these empirical realities into the data-derived CHAT-SD Framework, the study effectively bridges independent digital practice with formal classroom evaluation through an automated AI-Time Activity Segment embedded directly within the standard DepEd module flow. Operationalized via dual scannable QR gateway codes linked to customized, voice-enabled prompt threads and self-directed assessment scorecards, this structural intervention ensures that automated AI scores and instantaneous diagnostic feedback are never treated as final educational endpoints. Ultimately, this architecture establishes that digital readiness achieves its highest pedagogical utility when strictly governed as a controlled, preliminary sandbox engineered to act as a deliberate springboard that transfers well-prepared learners directly into active, unassisted, and face-to-face human interaction within the live classroom environment.

REFERENCES

1. Abdeljaoued, M. (2024). ChatGPT or the language genie: Perspectives and lived experiences of students in Tunisia. *Research Square*. https://doi.org/10.21203/rs.3.rs-5205322/v1
2. Ada, M. (2025). The attitudes of EFL learners towards the effectiveness of using ChatGPT to practise English grammar. *INTED Proceedings*, *1*, 6025. https://doi.org/10.21125/inted.2025.1557
3. Alghazo, R., Fatima, G., Malik, M., Abdelhamid, S. E., Jahanzaib, M., Nayab, D. E., & Raza, A. (2025). Exploring ChatGPT's role in higher education: Perspectives from Pakistani university students on academic integrity and ethical challenges. *Education Sciences*, *15*(2), Article 158. https://doi.org/10.3390/educsci15020158
4. Ali, Z., Bhar, S. K., Majid, S. N. A., & Masturi, S. Z. (2025). Exploring student beliefs: Does interaction with AI language tools correlate with perceived English learning improvements? *Education Sciences*, *15*(5), Article 522. https://doi.org/10.3390/educsci15050522
5. Andewi, W., Waziana, W., Wibisono, D., Putra, K. A., Hastomo, T., & Oktarin, I. B. (2025). From prompting to proficiency: A mixed-methods analysis of prompting with ChatGPT versus lecturer interaction in an EFL classroom. *Journal of English Language Teaching and Applied Linguistics*, *20*(2), 210–223. https://doi.org/10.64731/jsel.v20i2.282318
6. Annamalai, N., & Nasor, M. (2025). Exploring ChatGPT in education: Unveiling learners' experiences through the lens of self-determination theory. *Smart Learning Environments*, *12*(1), Article 8. https://doi.org/10.1186/s40561-025-00393-2
7. Ayman, S. E., El-Seoud, S. A., Nagaty, K., & Karam, O. H. (2023). The influence of ChatGPT on student learning and academic performance. *International Conference on Computer Applications*, *5*(1), 44–51. https://doi.org/10.1109/icca59364.2023.10401713
8. Azizah, R. N., Wijayanto, A., & Laila, M. (2025). Can AI-generated feedback overshadow teachers' existence? *Dinasti International Journal of Education Management and Social Science*, *6*(5), 3505–3518. https://doi.org/10.38035/dijemss.v6i5.4503
9. Balla, F. R. C., & Ramírez, E. A. V. (2025). Redefining speaking practice through ChatGPT's voice conversation: Towards autonomous and personalized practice. *Pacha Revista de Estudios Contemporáneos Del Sur Global*, *6*(18), Article e250430. https://doi.org/10.46652/pacha.v6i18.430

10. Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806
11. Capinding, A. T., & Dumayas, F. T. (2024). Transformative pedagogy in the digital age: Unraveling the impact of artificial intelligence on higher education students. *Problems of Education in the 21st Century*, *82*(5), 630–645. https://doi.org/10.33225/pec/24.82.630
12. Choi, W. C., & Chang, C. (2025). *ChatGPT-5 in education: New capabilities and opportunities for teaching and learning* (Preprint). MDPI Preprints. https://doi.org/10.20944/preprints202508.0684.v1
13. De, C. D. P., & Yee, C. P. (2025). Evaluating the impact of ChatGPT on academic performance and technology acceptance: A quasi-experimental study in Malaysian Chinese independent school history education. *International Journal of Academic Research in Business and Social Sciences*, *15*(8), 114–129. https://doi.org/10.6007/ijarbss/v15-i8/26115
14. Deep, P. D., Martirosyan, N. M., Ghosh, N., & Rahaman, M. S. (2025). *ChatGPT in ESL higher education: Enhancing writing, engagement, and learning outcomes* (Preprint). MDPI Preprints. https://doi.org/10.20944/preprints202503.1665.v1
15. Díaz-Suárez, V., Martín-Paciente, M., & Travieso, C. M. (2025). Exploring the impact of digital platforms on teaching practices: Insights into competence development and openness to active methodologies. *Applied System Innovation*, *8*(3), Article 64. https://doi.org/10.3390/asi8030064
16. Ding, D., & Yusof, A. M. (2025). Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: A mixed methods study. *Humanities and Social Sciences Communications*, *12*(1), Article 54. https://doi.org/10.1057/s41599-025-05550-z
17. Elmotri, B., Osman, W. A. E. A. E., Al-humari, M. A., Amri, F., Elyasa, Y. M., Malik, F. H., Garrouri, S., & Hassanein, O. H. S. (2025). Possibilities for improving ESP curriculum design and assessment strategies for Saudi universities with ChatGPT. *Arab World English Journal*, *1*, 234–249. https://doi.org/10.24093/awej/ai.13
18. Fleur, J. L. (2025). Multimodality and a diverse pedagogical mix: Insights into the 21st century South African secondary school learning environment. *Journal of Education*, *97*, 110–127. https://doi.org/10.17159/2520-9868/i97a06
19. Husin, N., Sedek, M., Sui, L. K. M., & Hakimi, H. (2025). Cross-curricular integration: Incorporating language competency skills in technical students' FYPs. *International Journal of Academic Research in Business and Social Sciences*, *15*(9), 443–456. https://doi.org/10.6007/ijarbss/v15-i9/26443
20. Jamshed, M., Albedah, F., Hussain, N., & Banu, S. (2025). Assessing the empowering efficacy and perceived pedagogical prospects of ChatGPT in the Saudi Arabian context: A mixed method study. *Journal of Language Teaching and Research*, *16*(3), 1002–1015. https://doi.org/10.17507/jltr.1603.31
21. Ji, H., Han, I., & Park, S. (2024). Teaching foreign language with conversational AI: Teacher-student-AI interaction. *Language Learning & Technology*, *28*(2), 91–112. https://doi.org/10.64152/10125/73573
22. Kabir, A. I., Ahmed, M. K., Begum, A., & Gomes, G. (2025). *Impact of AI technologies on academic integrity: Challenges, opportunities, and the plagiarism dilemma - Enhancing educational proficiency and credibility*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5177831
23. Karakuş, N., Gedik, K., & Kazazoğlu, S. (2025). Ethical decision-making in education: A comparative study of teachers and artificial intelligence in ethical dilemmas. *Behavioral Sciences*, *15*(4), Article 469. https://doi.org/10.3390/bs15040469
24. Khan, A., & Ann, O. W. (2025). The use of artificial intelligence (AI): ChatGPT and mastery of English-speaking skills. *International Journal of Research and Innovation in Social Science*, *9*(4), 6430–6442. https://doi.org/10.47772/ijriss.2025.90400465
25. Klar, M. (2025). Using ChatGPT is easy, using it effectively is tough? A mixed methods study on K-12 students' perceptions, interaction patterns, and support for learning with generative AI chatbots. *Smart Learning Environments*, *12*(1), Article 19. https://doi.org/10.1186/s40561-025-00385-2
26. Kuzminykh, I., Nawaz, T., Shenzhang, S., Ghita, B., Raphael, J., & Xiao, H. (2024). *Personalised feedback framework for online education programmes using generative AI*. arXiv. https://doi.org/10.48550/arxiv.2410.11904
27. Liao, M.-H. (2025). Cultivating proficient and efficacious L2 English speakers via VoiceThread-mediated self- and peer assessments. *Humanities and Social Sciences Communications*, *12*(1), Article 94. https://doi.org/10.1057/s41599-025-05674-2

28. Ma, M., Noordin, N., & Razali, A. B. (2025). Improving EFL speaking performance among undergraduate students with an AI-powered mobile app in after-class assignments: An empirical investigation. *Humanities and Social Sciences Communications*, *12*(1), Article 104. https://doi.org/10.1057/s41599-025-04688-0
29. Mahawan, A. (2025). Gender mainstreaming in ICT education: Exploring teachers' competence in selected academic institutions in Masbate, Philippines. *Pakistan Journal of Life and Social Sciences*, *23*(1), 492–504. https://doi.org/10.57239/pjlss-2025-23.1.00492
30. Maity, S., & Deroy, A. (2024). *Generative AI and its impact on personalized intelligent tutoring systems*. Open Science Framework. https://doi.org/10.35542/osf.io/kawr5
31. Masha, A. K., Mashologu, W., Baidoo, J., Kumanda, N., Makena, B., Kasumba, H., & Mutesasira, G. (2025). Supporting intermediate-phase mathematics teachers to improve learners' mathematics performance through continuous professional teacher development. *Journal of Education and Learning Technology*, *6*(5), 309–324. https://doi.org/10.38159/jelt.2025655
32. Meng, C. C., Fen, T. S., Ping, P. F., & Leong, A. (2025). Enhancing pre-service teachers' understanding of pedagogical key ideas by integrating ChatGPT in ODL course tutorials. *International Journal of Academic Research in Progressive Education and Development*, *14*(3), 114–129. https://doi.org/10.6007/ijarped/v14-i3/25773
33. Mohammed, S. J., & Khalid, M. (2025). Under the world of AI-generated feedback on writing: Mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia*, *15*(1), Article 12. https://doi.org/10.1186/s40468-025-00343-2
34. Murphy, T. H., Vaughn, G., Carpenter, R. E., McKinney, B., & McWhorter, R. R. (2024). *Examining the teacher readiness gap at the interface of artificial intelligence and medical education: A qualitative study of clinical educators*. Research Square. https://doi.org/10.21203/rs.3.rs-5362276/v1
35. Ng, B., Liu, J., Tong, X., Ye, K., Yenne, G., Chandrasekaran, V., & Li, J. (2025). *Analyzing security and privacy challenges in generative AI usage guidelines for higher education*. arXiv. https://doi.org/10.48550/arxiv.2506.20463
36. Ng, D. T. K., Leung, J. K. L., Su, J., Ng, C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, *71Block*(1), 137–161. https://doi.org/10.1007/s11423-023-10203-6
37. Nuñez, A. A. C., Nuñez, M. S. C., Pachay, J. F. Z., & Bosquez, A. M. C. B. C. (2025). Using ChatGPT Voice to improve speaking skills in English language learners. *Ciencia Latina Revista Científica Multidisciplinar*, *9*(1), 7143–7159. https://doi.org/10.37811/cl_rcm.v9i1.16390
38. Official Gazette. (2013). *Republic Act No. 10533: Enhanced Basic Education Act of 2013*. Government of the Philippines. https://www.officialgazette.gov.ph
39. Pagaduan, J. (2025). Artificial intelligence powered tools usage and writing proficiency of the junior high school students of Ambalayat Integrated School. *International Journal for Multidisciplinary Research*, *7Block*(3), Article 43394. https://doi.org/10.36948/ijfmr.2025.v07i03.43394
40. Park, S., Carlisle, D., Cheng, Z., Gillies, M., & Pan, X. (2025). Reducing foreign language anxiety through repeated exposure to a customizable VR public speaking application. *Frontiers in Virtual Reality*, *6*, Article 1519409. https://doi.org/10.3389/frvir.2025.1519409
41. Philippakos, Z. A. T., & Rocconi, L. (2025). AI literacy: Elementary and secondary teachers' use of AI-tools, reported confidence, and professional development needs. *Education Sciences*, *15*(9), Article 1186. https://doi.org/10.3390/educsci15091186
42. Phua, J. T. K., Neo, H. F., & Teo, C.-C. (2025). Evaluating the impact of artificial intelligence tools on enhancing student academic performance: Efficacy amidst security and privacy concerns. *Big Data and Cognitive Computing*, *9*(5), Article 131. https://doi.org/10.3390/bdcc9050131
43. Pohn, B., Mehnen, L., Fitzek, S., Choi, K., Braun, R. J., & Hatamikia, S. (2025). Integrating artificial intelligence into pre-clinical medical education: Challenges, opportunities, and recommendations. *Frontiers in Education*, *10*, Article 1570389. https://doi.org/10.3389/feduc.2025.1570389
44. Rahayu, A., Tarihoran, N., Rahmawati, E., Muslihah, E., Ma'mur, I., & Anita, A. (2025). Navigating the digital writing landscape: EFL students' perspectives on ChatGPT utilization. *International Journal of Learning Teaching and Educational Research*, *24*(5), 727–741. https://doi.org/10.26803/ijlter.24.5.37
45. Roe, J., & Perkins, M. (2024). *Generative AI in self-directed learning: A scoping review*. arXiv. https://doi.org/10.48550/arxiv.2411.07677

46. Sakaue, J., & Yamanaka, T. (2025). Enhancing English proficiency through AI conversations: The impact of ChatGPT on TOEIC and Speechace performance in a project-based learning context. *English Language Teaching*, *18*(3), 70–84. https://doi.org/10.5539/elt.v18n3p70
47. Shan, Z., Song, Z., Xu, J., Chen, W., & Chen, L. (2025). Complementing but not replacing: Comparing the impacts of GPT-4 and native-speaker interaction on Chinese L2 writing outcomes. *Behavioral Sciences*, *15*(4), Article 540. https://doi.org/10.3390/bs15040540
48. Syaripuddin, R. (2024). Enhancing EFL students' oral proficiency: The ChatGPT-assisted debate clinic method. *JELITA: Journal of English Language, Linguistics, and Literature*, *5*(2), 627–639. https://doi.org/10.56185/jelita.v5i2.651
49. Villaver, R. G., & Cabigas, M. E. A. (2025). Learning engagement with AI tools and academic performance. *International Journal of Multidisciplinary Research and Analysis*, *8*(7), 4191–4204. https://doi.org/10.47191/ijmra/v8-i07-55
50. Wang, X., Niu, J., Fang, B., Han, G., & Ju-hou, H. (2025). Empowering teachers' professional development with LLMs: An empirical study of developing teachers' competency for instructional design in blended learning. *Teaching and Teacher Education*, *165*, Article 105091. https://doi.org/10.1016/j.tate.2025.105091