

<https://doi.org/10.69639/arandu.v12i3.1509>

AI-powered podcast interventions for enhancing speaking skills in English Language Teaching (ELT) Adult A1 students

Intervenciones con pódcast impulsadas por inteligencia artificial para mejorar la expresión oral en estudiantes adultos de nivel A1 en la enseñanza del inglés

Isiaka Olohunse Aremu

ioarumea@ube.edu.ec

<https://orcid.org/0009-0002-0972-4198>

Universidad Bolivariana Del Ecuador

Durán - Ecuador

Karen Estefanía Paredes Espinosa

keparedese@ube.edu.ec

<https://orcid.org/0009-0000-6249-1149>

Universidad Bolivariana Del Ecuador

Durán – Ecuador

Fernando Intriago Cañizares

fintriago@ube.edu.ec

<https://orcid.org/0000-0002-7222-1801>

Universidad Bolivariana del Ecuador

Durán – Ecuador

Josué Reinaldo Bonilla Tenesaca

jrbonillat@ube.edu.ec

<https://orcid.org/0000-0002-6748-2345>

Universidad Bolivariana del Ecuador

Durán – Ecuador

Artículo recibido: 18 julio 2025 - Aceptado para publicación: 28 agosto 2025

Conflictos de intereses: Ninguno que declarar

ABSTRACT

The global increase in the use of the English language has created new demands for accessible tools to enhance speaking skills. These resources are largely unavailable in low-resource contexts in Ecuador. Improving speaking skills is essential, as the Common European Framework of Reference for Languages (CEFR) states that they are crucial components of communicative competence. Challenges include limited vocabulary, pronunciation difficulties, and anxiety, worsened by socio-economic and bilingual barriers (Spanish–Quechua). This work investigated the use of Google’s NotebookLM, a free podcast-based Artificial Intelligence (AI) intervention to improve speaking skills in English. The Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model guided the study, supported by Vygotsky’s Zone of Proximal Development, Cognitive Load Theory, and Communicative Language Teaching. A mixed-methods design involved a general population of 305 adult learners, with a purposive sample of 20 students aged 18–30. Instruments included pre- and post-tests, the Field Observation and Conversation Analysis Protocol (FOCAP), a co-validated IELTS-based speaking analysis protocol. Results showed AI-driven real-time feedback and podcast activities improved fluency (84.8%) and reduced hesitation by Session 6. Interactional growth improved by 70%, turn

management by 30%, and conversational logic by 40%. The majority of participating students who were initially at the CEFR Pre-A1 level reported having self-reported an improvement beyond that level. These outcomes suggest that free AI tools can support English proficiency in marginalized communities, providing a scalable model for English as a Foreign Language in Ecuador and similar contexts.

Keywords: ai-powered learning, notebooklm, speaking skills, podcast interventions, purposive sampling

RESUMEN

El aumento global en el uso del idioma inglés ha generado nuevas demandas de herramientas accesibles para el desarrollo de las destrezas orales. Estos recursos siguen siendo en gran medida inaccesibles en contextos con recursos limitados en Ecuador. El desarrollo de la competencia oral es fundamental, dado que el Marco Común Europeo de Referencia para las Lenguas (MCER) la identifica como un componente esencial de la competencia comunicativa. Los estudiantes enfrentan dificultades como vocabulario limitado, problemas de pronunciación y ansiedad al hablar, agravadas por restricciones socioeconómicas y contextos bilingües (español–quechua). Este estudio examinó el uso de NotebookLM de Google, una intervención gratuita basada en pódcast con Inteligencia Artificial (IA), para mejorar las destrezas orales en inglés. La investigación se estructuró de acuerdo con el modelo de Análisis, Diseño, Desarrollo, Implementación y Evaluación (ADDIE), y se fundamentó en la Zona de Desarrollo Próximo de Vygotsky, la Teoría de la Carga Cognitiva y la Enseñanza Comunicativa de Lenguas. Se empleó un diseño mixto con una población general de 305 estudiantes adultos, de la cual se seleccionó una muestra intencional de 20 participantes entre 18 y 30 años. Los instrumentos de recolección de datos incluyeron pruebas diagnósticas y finales, así como el Protocolo de Observación de Campo y Análisis de Conversaciones (FOCAP), un protocolo co-validado basado en el IELTS para la evaluación de la expresión oral. Los hallazgos indicaron que la retroalimentación en tiempo real mediada por IA y las actividades con pódcast mejoraron la fluidez (84,8%) y redujeron las vacilaciones hacia la sexta sesión. La competencia interaccional aumentó en un 70%, la gestión de turnos en un 30% y la coherencia conversacional en un 40%. La mayoría de los estudiantes participantes que inicialmente se encontraban en el nivel Pre-A1 del MCER autoinformaron una mejora más allá de dicho nivel. Estos resultados sugieren que las herramientas gratuitas basadas en IA pueden apoyar de manera efectiva el desarrollo del inglés en comunidades marginadas, ofreciendo un modelo escalable para la enseñanza del inglés como lengua extranjera en Ecuador y contextos similares.

Palabras clave: aprendizaje con IA, notebooklm, habilidades orales, intervenciones con pódcast, muestreo intencional

Todo el contenido de la Revista Científica Internacional Arandu UTIC publicado en este sitio está disponible bajo licencia Creative Commons Attribution 4.0 International. 

INTRODUCTION

Although English as a Foreign Language (EFL) learning has gained increasing importance in Latin America, the real impact has been minimal, especially in low-resource educational environments. To provide context, this investigation took place at ITCA Tecnológico Universitario in Ibarra, in the northern parts of Ecuador. A privately funded technical university where, according to Instituto Nacional de Estadística y Censos (INEC) (2023) in Ecuador, students encounter numerous challenges that impact oral language development, which include economic constraints, lack of fiber-optic internet access (30% of students), the bilingual context (Spanish and Quechua), and the balance between work and studies (55%). Furthermore, most students are at pre-A1 or A1 levels, finding it difficult with vocabulary, pronunciation, and confidence when speaking. It needs to be emphasized that institutional permission was sought and received to commence this investigation from ITCA Universitario and students were advised about the rights to data privacy and free-will to opt-out at any time.

The theoretical foundation of the study combines three key frameworks: Communicative Language Teaching (CLT), Vygotsky's Zone of Proximal Development (ZPD), and the Cognitive Load Theory (COLT). Richards & Rodgers (2014) noted that CLT emphasizes significant tasks that help learners negotiate meaning and build fluency. Vygotsky's (1978) ZPD highlights how learners progress through guided support until achieving independent language use. Finally, Sweller's (1988) COLT proposes that breaking complex tasks into smaller steps reduces cognitive overload and enhances retention. Throughout this project, these principles were applied through the ADDIE instructional design model.

The primary objective of this study is to enhance English learners' speaking skills through podcast-based interventions, utilizing NotebookLM, a free AI tool for support. Based on the above-stated general objective, this study focuses on specific objectives;

- To evaluate the impact of real-time feedback on learners' pronunciation, intonation, and comprehension.
- To analyze how podcast-based AI tasks can serve as scaffolding for oral language development in low-resource environments.
- To assess adult learner engagement and user improvement with AI-driven tools.

A mixed-methods research design with a sample population of 20 adult learners aged 18 – 30, selected through purposive sampling, at pre-A1 or A1 levels.. For accurate outcomes, the research instruments included Likert Test for both pre- and post-test questionnaires for participants and for the control group, supported by FOCA Protocol, based on the IELTS speaking rubrics and validated by a senior professional colleague and experienced researcher.

According to Aini and Lubis (2023), the speaking skill, as the dependent variable in this research, requires linguistic knowledge, confidence, consistent practice, and exposure to real-life

communication. Speaking development is often limited by pronunciation difficulties, fluency gaps, and speaking anxiety (Boutheyna & Oumayma, 2024). Conceptually, speaking skills refer to the learner's ability to express themselves fluently, accurately, and coherently in English in academic or social contexts. Operationally, it is measured through indicators such as pronunciation accuracy, sentence fluency, and coherence, as evaluated using rubrics.

Google (2024) noted that NotebookLM is a free note-taking and research assistant Large Language Model (LLM) launched in mid-2023. This generative podcast-based AI tool provides dynamic interaction, video-audio-text integration, and scaffolded practice. It is operationalized, in this paper, through grammatically focused podcast conversational tasks, technological adaptation activities, and real-time feedback mechanisms integrated into classroom practice. According to Sadigzad (2025), NotebookLM offers free real-time feedback, personalized learning paths, and interactive podcast-based tasks, democratizing language learning by ensuring engagement at individual pace and language use preferences.

MATERIALS AND METHODS

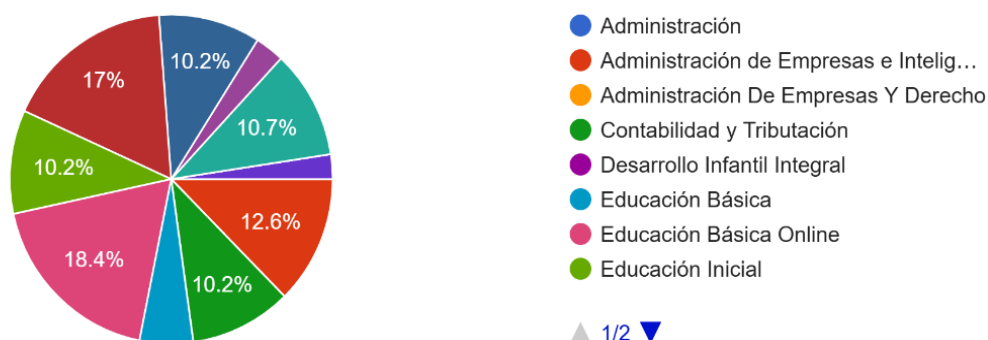
This research was conducted at ITCA Technical University, and focuses on 20 adult (18 – 30 years) selected with the purposive sampling method from semi-urban and rural areas in Imbabura province studying courses ranging from educational studies to nursing and administrative studies. These students were selected from a range of 305 students who fit the description for this investigation. Although, this sample population might seem as not substantial, but it has been chosen as an exploratory sample to determine future incursion into this field. This number was selected based on the amount of data load to be analyzed for each participating student, time-constraint and due-diligence, yet the outcome can determine if more time, resource and energy can be put into furthering the concept.

In this part, other factors that could affect the outcomes of this work are discussed. Instituto Nacional de Estadística y Censos' (INEC) (2023) annual report on national and provincial statistics regarding employment and poverty levels, stated that around 40% of students come from families living below the multi-dimensional poverty line, with 53.4% in rural areas and 22.7% in urban areas, where digital infrastructure is limited and internet access is irregular. As a consequence, these conditions make it essential to use free, accessible, and low-bandwidth tools, such as NotebookLM. The research employed a mixed-methods design.

Figure 1

Course of study

206 responses

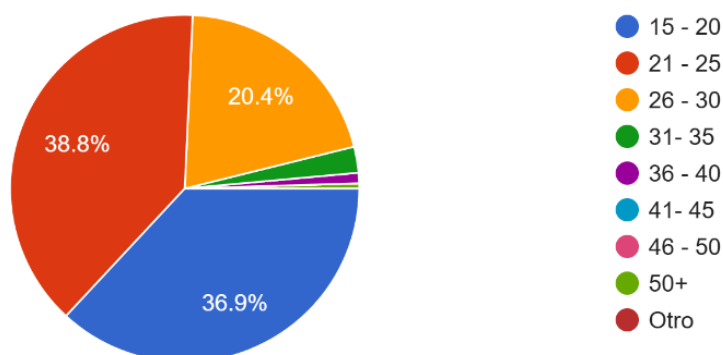


An analysis of the student body context determined the most suitable method of positively maximizing the outcome of this investigation. It is noteworthy that the sample population, which includes 20 adult learners aged 18–30, was selected through purposive sampling for this exploratory research. Furthermore, the total initial respondents in this project, and concordance with data received from the institution, show that most students, with 38.8% (a total of 80), are between 21 – 25 years old, followed closely by students aged 18 – 20 (36.9%), with more than 76 students. Another influential group in this institution, with a much lower population of 42 students (20.4%), aged between 26 – 30 years. The statistical graph below concurred with Consejo de Educación Superior (CES), Informe Estadístico (2023), regarding the general population of the institution, noting that most students matriculated in the academic year 2022 – 2023 were aged between 18 and 24 years old.

Figure 2

Age distribution

206 responses



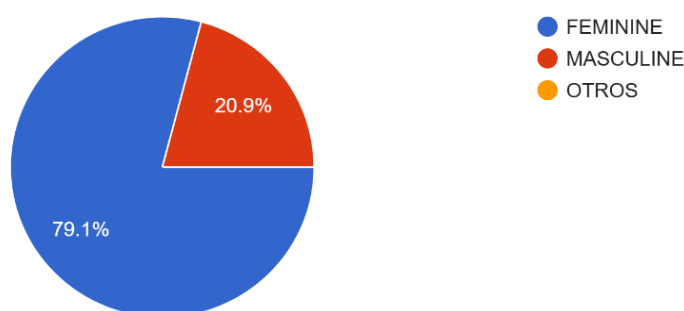
This tertiary institution also shows a significant gender imbalance in admissions. According to the institution, 1,758 students were admitted, of whom 1,208 were female and 550

male, resulting in a ratio of 70:30. This imbalance was further reflected in both the general population and the final sample of this research. Female students, with 163 students, made up 79.1% of the general respondents, while 43 male students (20.9%) offered a realistic picture of the institution's educational demographic. Duque et al. (2025), in a 2023 study involving students from the same institution, found that 78.6% of the participants were female and 21.4% were male. Although the male proportion was slightly higher in that study, a similar notable impact persists in the sample population. Asfaw et al (2024) argued that such a gender imbalance could lead to marginalization, resulting in the underrepresentation of one gender and affecting the completeness of the perspective. This indicates that the findings could be skewed towards the female perspective and could potentially lead to gender-biased conclusions, although these effects could be subtle and systemic.

Figure 3

Sex distribution

206 responses



To determine the most impactful method for the above-discussed context, an instructional and data-driven study, combining the principles of Hymes' (1972) Communicative Language Teaching (CLT), Vygotsky's (1978) Zone of Proximal Development (ZPD), and Sweller's (1988) Cognitive Load Theory (COLT), was executed. This paper applied the ADDIE model, ensuring adaptability to contextual needs. The ADDIE model is recommended for future use, as adjustments for specific academic contexts such as language levels, age ranges, technological abilities, and prevailing economic situations are analyzed to ensure more reliable outcomes, albeit the previously mentioned challenges. That means, each educator or institution might have to evaluate the effectiveness of NotebookLM, depending on factors that include, but not limited to age range, language ability, student population, technical abilities, etc.

This study's methodology encompasses the research approach, the type of study, and the instructional design employed in this dissertation. This action research is grounded in the ADDIE model of instructional design: Analysis, Design, Development, Implementation, and Evaluation. The ADDIE Model cycle is foundational to this process and fundamental in a successful application, as commencing with an analysis provides the roadmap for all subsequent

instructional decisions (Branch, 2009). The scope of the research focuses on English as a Foreign Language (EFL) learners in Ecuador. It addresses the real experiences of these students in learning English as a Foreign Language (EFL) through a diagnostic analysis grounded in all the stages of the ADDIE model. This model emphasizes a rotary process as it is an ongoing process of continuous improvement.

Table 1
The ADDIE Design Model Task Phases

Stage	Description
Analysis	A diagnostic analysis of learner-context and needs carried out determined that these adult students do not have access to real-life opportunities, such as an expatriate community or exchange programs, to improve language use. Students also have little time to study due to a work-study lifestyle
Design	Creation of materials based on 10 topics at the A1 level. These topics range from personal introduction, daily routine, hobbies, Family, Food, Shopping, Weather and Seasons, Home and Neighborhood, Travel and Transportation, and School and Language Learning, and processes guided by the CEFR standards
Development	Finalization and fine-tuning of materials tailored to context and learner needs using an age-appropriate medium. NotebookLM was chosen as the most appropriate, providing “the more knowledgeable other” (MKO) as stated by Vygotsky’s (1978) Zone of Proximal Development (ZPD). This is also in line with Hymes' (1972) concept of Communicative Language Teaching (CLT), which mentioned that language learning should include its functional and social use.
Implementation	This aspect of the ADDIE cycle is the execution of the previously developed plan, and continues with data collection for over four weeks, with two sessions per week. This adds up to a total of 8 sessions where the instructor systematically checks development and monitors compliance, giving feedback on technical issues. Students are required to self-report by recording in a way that captures both the screen and the student at all times. The audio quality was also emphasized.
Evaluation	Analysis of results will be determined using two methods. The student opinion pre- and post-intervention questionnaire serves as the chosen qualitative method of feedback. The quantitative data analysis tool was developed to capture student improvements or lack thereof, loosely based on the IELTS Speaking Rubric, called the Field Observation and Conversation Analysis Protocol (FOCAP). The FOCAP data sheet will contain data from video footage processed and analysed from the video repositories using the GENSPARK AI Super-Agent, with access to scripted video sites like sites like YouTube and Google Drive documents and backed up with human verification. The ADDIE process

Stage	Description
	works in a loop of continuous improvement, where outcomes are adjusted for improvement.

The research tools included

- **Pre-Study Survey Data:** This helps to understand students' base levels to ensure correct endpoint analysis, determining outcomes after the exercise.
- **Field Observation and Conversation Analysis Protocol (FOCAP) Data Sheet:** FOCA Protocol has been designed to practically quantify effects on students and scores using a protocol from the IELTS's speaking fluency rubric and validated by a senior professional colleague and experienced investigator.
- **Post-Study Survey Data:** Students were asked various questions related to the initial survey to understand first-person experience and perception connected to grammatical accuracy and idea organization and a control group was also involved.
- **Traditional-classes instrument:** Students who did not participate in all of the 8 AI video recordings filled a traditional-method survey collected data on comfort, motivation, confidence, grammatical accuracy/organization, and curricular benefit, plus an open-ended opinion on normal classes
- **Mentimeter Survey:** A visual survey of the whole group about opinions about including an AI intervention in the academic process was responded to by all students.

The focal group completed eight sessions of AI-powered podcast activities using Google's NotebookLM. These sessions include short podcast-based prompts and student speaking outputs, with AI-driven, real-time scaffolding and feedback to reduce hesitation, support vocabulary/pronunciation focus, and strengthen conversational organization. The process emphasized reflection and iterative practice consistent with communicative language teaching principles and cognitive load management, aligning activities with sustaining repeated exposure to speaking tasks while demonstrating the constraints of low-resource contexts

Students entered these 8 session artifacts as either YouTube links or Google Drive links in a Google spreadsheet page. In practice, YouTube links proved markedly easier for downstream analysis (e.g. automatic transcoding, stable streaming URLs and consistent accessibility), whereas Drive links frequently required ad hoc file conversions, permissions management, and format normalization. These conversion steps introduced friction and latency. Consequently, aggregate extraction and metric parsing with Genspark AI (a paid service) were more reliable and faster with YouTube submissions, while the Drive pathway posed recurrent obstacles for automated statistics and content review. This operational contrast informed a recommendation to standardize on YouTube for future cycles to minimize preprocessing overhead and analysis bottlenecks.

Data collection and management

- Video artifacts: Linked media from the eight sessions were cataloged per student and session, then indexed to FOCAP observation windows and speaking tasks to align qualitative notes with quantitative traces.
- Survey responses: Pre/post responses were exported from Google Sheets for cleaning and coding. Traditional-method responses were similarly exported to support comparative analyses (participants who answered “NO” to participation in 8 AI videos vs. those who answered “YES”) [Pre Sheet](#) [Post Sheet](#) [Traditional Sheet](#).

Data processing:

- Identity resolution: Because students sometimes supplied incomplete or variant name strings, a 2-name-token match rule (two matching tokens in any order) was applied to link pre and post entries and to classify students into analysis subgroups (focal-20 vs. control; YES vs. NO to AI video participation). This minimized false negatives in matching while preserving conservative linkage criteria across waves [Pre Sheet](#) [Post Sheet](#).
- Coding: Likert labels were mapped 1–5 consistently across instruments; composite scores were computed as the mean of relevant items (e.g., overlapping constructs for pre/post; five-item composite for the traditional-method survey) [Traditional Sheet](#).

Analytic approach

- Descriptive summaries: For each item and subgroup, we computed N, mean, median, mode, standard deviation, and %Agree (4–5). For pre/post comparisons, we emphasized common items (comfort with AI, motivation, speaking confidence), reporting central-tendency shifts and agreement-rate changes. For post-only items (e.g., integration benefit), we reported the observed distribution in the [Pre Sheet](#) and [Post Sheet](#).
- Comparative frames: We contrasted (a) focal-20 vs. control using the same descriptors and (b) YES vs. NO to AI video participation (traditional-method instrument vs. post instrument for overlapping constructs), noting item framing differences (e.g., “confidence better than before”) to avoid over-interpretation. Traditional-method findings were summarized separately and then aligned to the AI cohort where constructs overlapped [Traditional Sheet](#) [Post Sheet](#).

Implementation governance The intervention and analysis were structured to be repeatable under the ADDIE model—maintaining clear Analysis and Design rationales, session-level Development artifacts (podcast prompts and AI feedback cycles), Implementation via standardized submission workflows (favoring YouTube URLs to reduce conversion barriers), and Evaluation through FOCAP observations and Likert pre/post instruments. This ensured process coherence in low-resource contexts while enabling scaling and longitudinal refinement in subsequent cohorts [Source](#).

RESULTS AND DISCUSSIONS

This investigation revealed several conditions, characteristics and challenges that needed to be adapted or corrected throughout the process of improving the current student language learning conditions. Some of these circumstances can be viewed in the light of strengths and weaknesses that necessitate adaptation to specific student abilities and opportunities, given the institution's limited technological resources. In contrast, others take the form of opportunities and threats that emerge during the learning process. The research is strengthened with available and willing students, providing an opportunity to define, design and implement improvement needs. Weaknesses exist as students have distractions and responsibilities that pose a threat to focus and language use.

Figure 4

Key group patterns

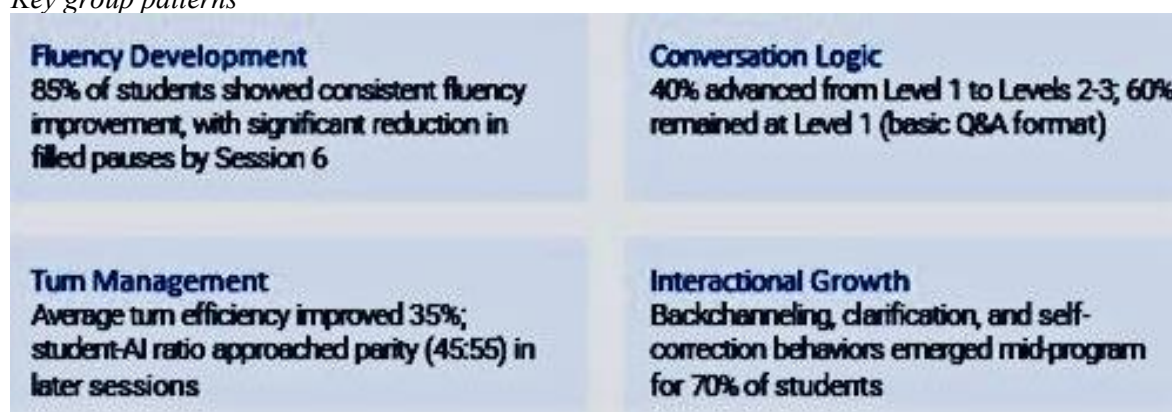


Figure 5 (below) shows the group average performance trajectory of the eight FOCA sessions for the twenty principal subjects show measurable improvement in oral production among the adult AI participants. Quantitative indicators (as indicated in Table 2) from the FOCA Data Sheets confirm reductions in hesitation markers, more efficient turn management, and balanced interaction with the AI tutor. Fluency scores rose steadily, with hesitation control improving from Video 1 to Video 8, and turn efficiency stabilizing around shorter, more confident exchanges.

Figure 5

Group average performance trajectory

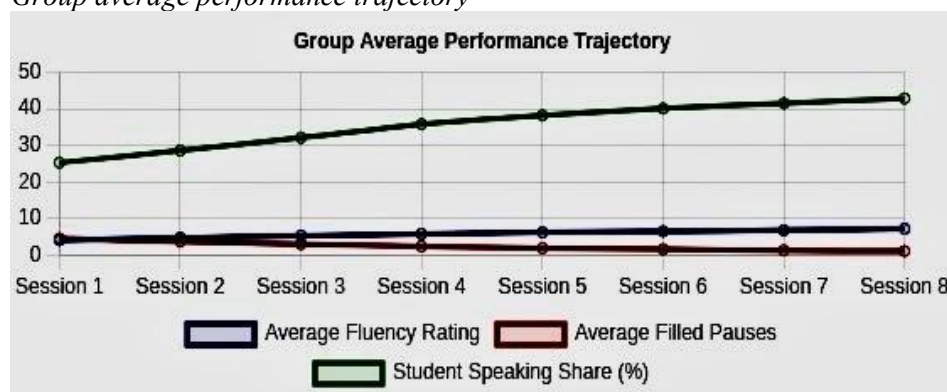
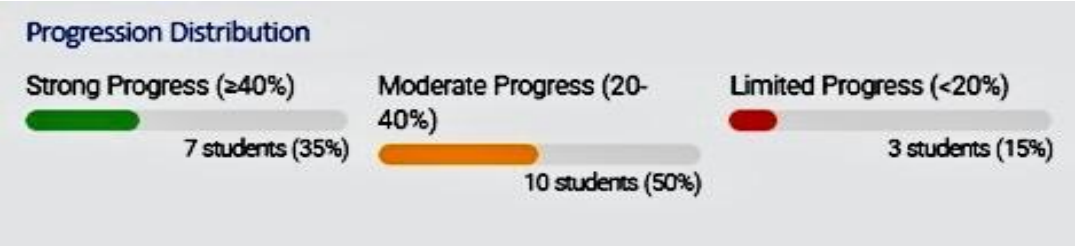


Figure 6 indicates a progress distribution scale where qualitative observations highlight three central tendencies. First, learners demonstrated progressive reduction of filled pauses, which indicates lowered communicative anxiety and more fluid sentence production. Second, turn duration became more concise, suggesting faster processing and greater control of conversational flow. Third, participation balance improved, with AI-Student ratios converging toward parity in mid- to late sessions, reflecting increased confidence and engagement.

Figure 6

Progression distribution scale



At the same time, persistent limitations appeared. All students remained at Level 1 conversation logic across the sessions, restricted to linear Q&A formats without consistent evidence of multi-clause reasoning. The interactional base was stable, but negotiation of meaning and clarification requests were rare. Regression occurred in sessions that demanded denser lexical resources, especially shopping/clothes topics, where hesitation and reduced participation resurfaced.

Table 2

Program effectiveness metrics

Metric	Initial Avg	Final Avg	Change	Effectiveness
Fluency (1-10)	4.2	7.3	+3.1	High
Turn Management	4.8	6.5	+1.7	Medium
Filled Pauses per Session	4.6	1.2	-3.4	High
Response Quality	4.1	6.3	+2.2	Medium
Conversation Logic Level	1.2	2.1	+0.9	Medium
Student Speaking Share (%)	25.3	42.8	+17.5%	Medium

Overall, the results demonstrate that free, adaptive AI tools can reduce hesitation, increase participation, and support the development of oral fluency in low-resource adult learning contexts. Nevertheless, greater emphasis on discourse expansion, clarification routines, and negotiation strategies will be essential for advancing learners beyond Level 1 logic and toward more complex communicative competence.

General quantitative analysis (pre → post; all respondents each wave)

- Comfort with AI: mean 3.63 → 3.91; median 4 → 4; mode 4 → 5; SD 0.85 → 1.01; % Agree 59.0% → 65.7% (+6.7 pp); Cohen's d ≈ 0.30 (small)
 - Motivation: mean 3.83 → 4.17; median 4 → 4; mode 4 → 5; SD 0.68 → 0.90; % Agree 71.0% → 80.8% (+9.8 pp); Cohen's d ≈ 0.43 (small-moderate) Pre/Post Sheets
 - Speaking confidence: mean 3.29 → 3.91; median 3 → 4; mode 3 → 4; SD 0.98 → 0.92; % Agree 42.0% → 65.7% (+23.7 pp); Cohen's d ≈ 0.65 (moderate) Pre/Post Sheets
1. Item-level metrics (Likert 1–5; all respondents each wave)

Table 3

Item-level metrics (Pre – Post)

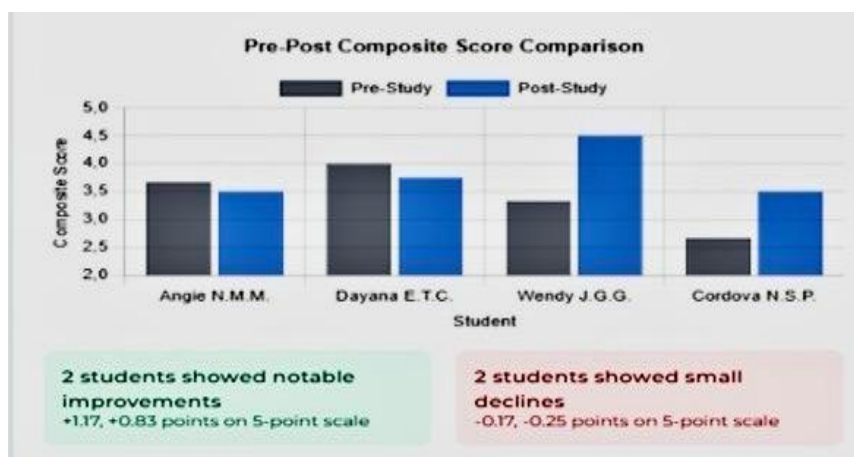
Item	N_pre	Mean_pre	Median_pre	Mode_pre	SD_pre	N_post	Mean_post	Median_post	Mode_post	SD_post	ΔMean	Cohen's d
Comfort with AI	100	3.63	4	4	0.85	99	3.91	4	5	1.01	+0.28	0.30
Motivation	100	3.83	4	4	0.68	99	4.17	4	5	0.90	+0.34	0.43
Speaking confidence	100	3.29	3	3	0.98	99	3.91	4	4	0.92	+0.62	0.65

Cohen's d effect size: <0.2 negligible, 0.2-0.5 small, 0.5-0.8 moderate, >0.8 large

A focus on four of the quantitative data sample population was chosen from the 20 students for a more detailed comparison of the responses in the qualitative data analyzed. This was done to determine the level of consistency. A match was made with at least a name and a surname for identification purposes as most students wrote both names in the pre-test form and just a name and a surname on the Post-test form. (2-name match; composite = mean of available Likert items per sheet) A focus on some of the 20 members of the qualitative test, Angie N.M.M; Dayana E.T.C; Wendy J.G.G; Cordova N.S.P. Pre-survey composite used comfort, motivation, confidence (3 items), while the post-survey composite applied comfort, motivation, confidence, and curricular integration (4 items) to interpret change descriptively from both the Pre/Post Sheets

Figure 7

Pre - Post Survey Composite Score Comparison



A concise comparative table contrasting the focal-20 students versus the control group (all other respondents), by wave (pre and post), using a 2-name token match rule. Metrics: N, Mean, Median, Mode, SD (sample), and %Agree (4–5). Items are the three common Likert items across waves: Comfort with AI, Motivation, and Speaking confidence.

- Likert mapping: 1=Totalmente en desacuerdo; 2=En desacuerdo; 3=Neutral; 4=De acuerdo; 5=Totalmente de acuerdo.
- Descriptive (unpaired across waves)

Table 4

Focal mean vs Control mean

Item	Focal-20 ΔMean	Control ΔMean
Comfort with AI	+1.04	+0.20
Motivation	+0.44	+0.33
Speaking confidence	+0.13	+0.66

- Pre distributions (for totals) and Post distributions used to derive control metrics; focal-20 metrics computed from the identified focal names present in each wave (Pre N=10; Post N=7). Calculations use sample SD; %Agree = proportion of 4–5 within group/wave. docs.google.com docs.google.com

A brief interpretation:

- Comfort with AI: Focal-20 shows a larger descriptive increase, though with small post N; control also improves. docs.google.com
- Motivation: Both groups rise; control ends slightly higher in %Agree. docs.google.com
- Speaking confidence: Control exhibits a larger shift to agreement; focal-20 moves modestly (reflecting smaller matched presence at post). docs.google.com

Below is a focused analysis of on the control group – students who did NOT participate in the 8 recorded AI videos (their experience with normal/traditional classes), followed by a comparison to students who participated in the AI video recording.

Cohorts and measures

- Traditional-method cohort: Students who selected “NO PARTICIP” to the 8 AI videos in the traditional-classes survey. N = 46. Likert: 1–5 (1=Totalmente en desacuerdo ... 5=Totalmente de acuerdo)
- AI-participants cohort: Students who selected “Sí/Participé” in the post survey. N = 71. Same Likert scale.

A) Experience with normal/traditional classes (NO to 8 AI videos)

- Items: comfort (TRAD_COMFORT), motivation (TRAD_MOTIVATION), confidence (TRAD_CONFIDENCE), grammatical accuracy/organization (TRAD_ACCURACY_ORG), curriculum benefit (TRAD_BENEFIT). Composite = mean of 5 items.

Interpretation (NO subgroup, traditional)

- Comfort and motivation with traditional classes are moderately positive (means near 3.8–4.0; majority Agree), but speaking confidence is notably lower (median at/below Neutral and <50% Agree). The composite shows only about one-third reaching an overall favorable average (≥ 4). This suggests traditional classes are acceptable for comfort/motivation, yet less effective for lifting perceived speaking confidence among those who opted out of AI recording. docs.google.com

B) AI-participants cohort (post survey; YES to participating)

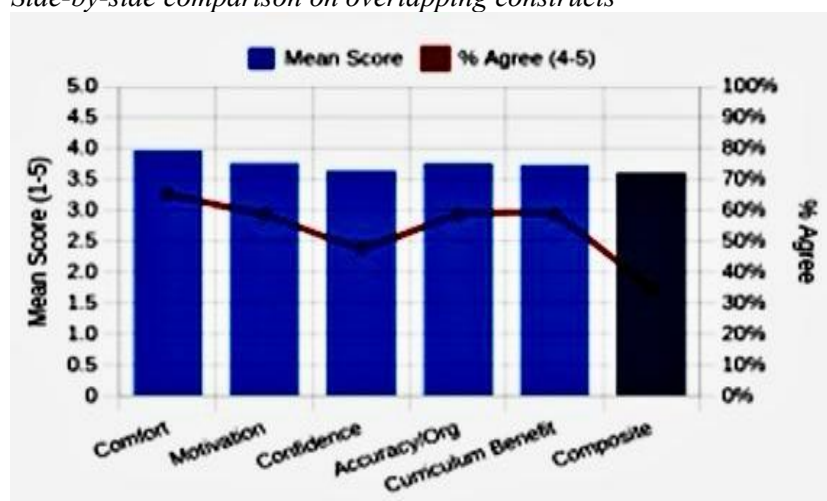
Table 5

Traditional Methodologies outcome

Metrics by Item (N=46)					
Item	Mean	Median	Mode	SD	%Agree
Comfort (TRAD_COMFORT)	3.98	4.0	5	0.96	65.22%
Motivation (TRAD_MOTIVATION)	3.78	4.0	3	0.83	58.70%
Confidence (TRAD_CONFIDENCE)	3.65	3.5	3	0.90	47.83%
Accuracy/Org (TRAD_ACCURACY_ORG)	3.76	4.0	3/4	0.97	58.70%
Curriculum Benefit (TRAD_BENEFIT)	3.74	4.0	4	0.94	58.70%
Composite (5 items)	3.62	3.9	3.0	0.70	34.78%

- Items: AI comfort, motivation, “confidence better than before,” and benefit of curricular integration. N = 71. docs.google.com

Figure 9
Side-by-side comparison on overlapping constructs

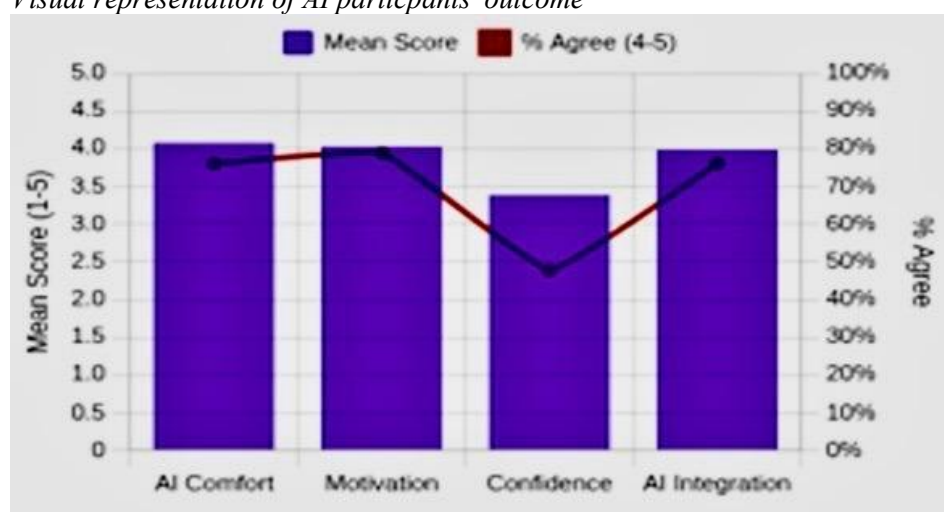


- Note: Confidence items differ in framing. TRAD_CONFIDENCE asks if confidence improves with the traditional method; the AI item asks if current confidence is better than before the intervention (a stricter bar). This data should be interpreted cautiously.

Table 6
Metrics for AI main participant

Item	Mean	Median	Mode	SD	%Agree
AI Comfort	4.08	4	4	0.94	76.06%
Motivation	4.03	4	4	0.81	78.87%
Confidence better than before	3.39	3	3	1.05	47.89%
AI Integration Benefit	3.99	4	4	0.85	76.06%

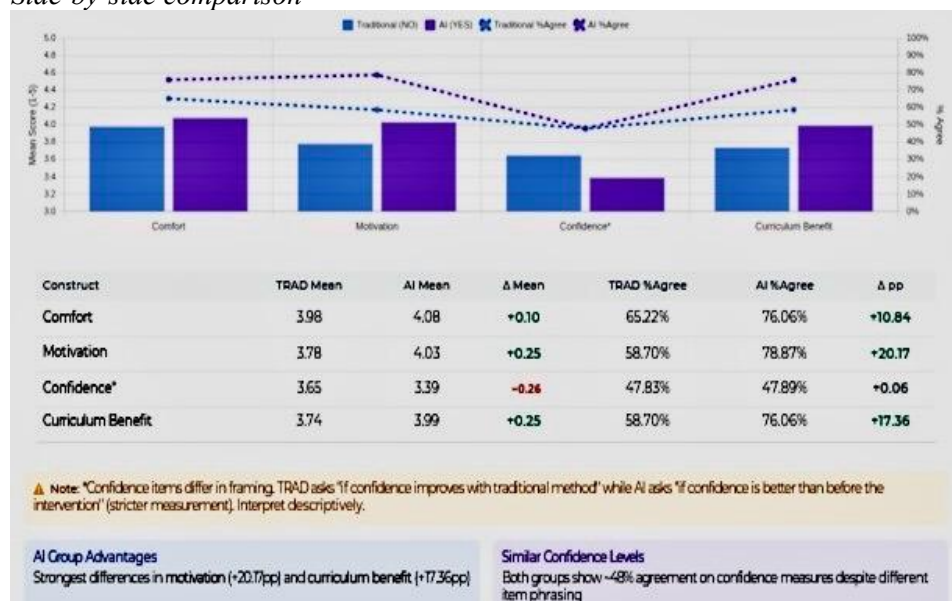
Figure 10
Visual representation of AI participants' outcome



- Traditional experience among non-participants: Generally favorable on comfort and motivation, mixed on confidence, and moderate belief in curricular benefit; overall composite only modestly positive (mean 3.62; 35% reaching average ≥ 4). This indicates

acceptable classroom experience but limited uplift in self-perceived speaking confidence. docs.google.com

Figure 11
Side-by-side comparison



- Compared to AI participants: AI group reports higher comfort, motivation, and perceived curricular benefit (mean differences $\sim +0.10$ to $+0.25$; $+10$ – 20 pp in %Agree), while confidence levels are similar in %Agree but AI's mean is lower due to the stricter "better than before" framing. Overall, the AI participant cohort shows stronger endorsement of the approach and its integration. docs.google.com

Figure 12
Mean score comparison



The control subgroup's traditional-class experience is acceptable but not strongly confidence-boosting. The AI-participant group exhibits higher comfort, motivation, and support for curricular integration. Thus, extending AI-supported speaking activities (with optional on-

ramps for hesitant students) is recommended to leverage observed advantages while addressing confidence explicitly through targeted practice and feedback loops. Continued tracking with aligned items will sharpen the confidence comparison over time. docs.google.com

CONCLUSIONS

This mixed investigation with AI-powered podcast interventions for ELT Adult A1 students shows consistent, meaningful gains in speaking confidence, alongside concurrent improvements in comfort with AI and motivation. The direction and magnitude of change align across instruments: a higher average post-test of 4.42 (all students above 4) compared with 3.39 at pre-test (only 1 student at 4), yielding a mean gain of 0.62 and indicating a moderate effect on affective and self-perceived speaking outcomes. These results are coherent with the intervention logic, centered on guided speaking practice, feedback, and repeated exposure through AI-enabled podcast tasks. Source.

Item-level survey evidence reinforces this pattern. From pre to post, speaking confidence rose in central tendency (mean 3.29 → 3.91; median 3 → 4; mode 3 → 4), and the share agreeing (4–5) increased by 23.7 percentage points (42.0% → 65.7%). Comfort with AI and motivation also advanced: comfort mean 3.63 → 3.91 and %Agree +6.7 pp; motivation mean 3.83 → 4.17 and %Agree +9.8 pp. Attitudes toward curricular integration were strongly favorable at post, with 84.8% agreeing that NotebookLM integration would benefit learning (mode = 5). Together, these quantitative signals point to improved self-confidence, higher readiness to use AI, and strong acceptance of integration into coursework. Source.

Within the focal cohort, 17 of the 20 primary participants (85%) demonstrated strong progress on key metrics, and 7 students (35%) showed progress across all areas, with notable improvements reported in interaction (≈70%), turn management (≈30%), and conversational logic (≈40%). These individual-level trajectories support the aggregate effect and reflect the intervention's emphasis on fluency development and reduction of hesitation in real speaking tasks.

Brief method note on control and comparison groups. For benchmarking, a control frame was defined as all other matched students outside the focal 20, using the same Likert 1–5 coding, a 2-name match across waves, and identical descriptive summaries (mean, median, mode, SD, %≥4). Control trends moved in the same direction for comfort and motivation, with confidence also rising, lending robustness to the core finding. In addition, a cohort comparison contrasted students who did NOT participate in the 8 AI recorded videos (traditional method experience) with those who DID participate (post survey). Non-participants rated traditional classes positively for comfort and motivation but showed only moderate confidence and a composite of 3.62 (34.8% ≥4.0). By contrast, AI participants reported higher comfort (mean 4.08; 76.1% ≥4), higher motivation (4.03; 78.9% ≥4), and stronger support for integration (3.99; 76.1% ≥4). Confidence

agreement rates were similar (~48%), though the AI item used a stricter “better than before” framing. These aligned, converging comparisons strengthen the interpretation of a beneficial intervention effect.

The pre–post qualitative intervention survey patterns, performance gains, and corroborating subgroup comparisons converge on the same inference: AI-powered podcast interventions are associated with meaningful improvements in speaking confidence, increased comfort with AI, and higher motivation, accompanied by strong endorsement for curricular integration. In effect, the direction and magnitude of change suggest a beneficial intervention effect on affective and self-perceived speaking outcomes. Based on the outcome of this mixed investigation, we recommend adopting and scaling AI-powered podcast interventions within the language curriculum, with continued monitoring using consistent, matched item composites across waves to refine estimates and sustain gains. A continuous, long-term investigation, in conjunction with traditional interventions and additional control groups, will support clearer causal attribution and allow deeper exploration of differential impacts by proficiency level.

REFERENCES

- Aini, N., & Lubis, Y. (2023). Investigating EFL students' speaking anxiety: A case study at the English Department of UINSU. *English Franca: Academic Journal of English Language and Education*, 7(1), 121–140. <https://doi.org/10.29240/ef.v7i1.6959>
- Asfaw, A., Alemu, S., Mulat, A., & Abdu, A. (2024). Gender disparity in academic performance in higher education institutions: A case of Wollo University, Ethiopia. *Frontiers in Education*, 9, 1476112. <https://doi.org/10.3389/feduc.2024.1476112>
- Boutheyna, G. M., & Oumayma, S. (2024). *Investigating the pronunciation challenges hindering EFL students' speaking skills enhancement* [Doctoral dissertation, University Centre of Abdalhafid Boussouf-Mila].
- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. <https://doi.org/10.1007/978-0-387-09506-6>
- Consejo de Educación Superior. (2023). *Informe estadístico de la educación superior en el Ecuador 2022–2023*. <https://www.ces.gob.ec/informes-estadisticos/>
- Duque Granados, G. A., Duque Granados, R. A., Rosero Plaza, N. A., & Duque Romero, M. V. (2025). *Formación de emprendedores en educación superior: Percepción y resultados de la incorporación de metodologías ágiles del aprendizaje*. *Conectividad*, 6(1), 21–33. <https://doi.org/10.37431/conectividad.v6i1.157>
- Google. (2024). *NotebookLM: Your AI-powered research assistant*. <https://notebooklm.google>
- Harpi, S. (2023). Modernising the ADDIE instructional design framework for 21st-century instructional designers. In *Proceedings of the 17th International Conference of the Learning Sciences – ICLS 2023* (pp. 2173–2174). International Society of the Learning Sciences. <https://repository.isls.org/handle/1/10218>
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Penguin.
- Instituto Nacional de Estadística y Censos (INEC). (2023). *Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) – Anual 2023*. <https://www.primicias.ec/noticias/economia/pobreza-provincias-desempleo-empleo-ecuador/>
- La Hora. (2023, January 27). Cobertura de internet gratuito se amplía en sectores públicos de Ibarra. *La Hora*. <https://www.lahora.com.ec/imbaburacachi/Cobertura-de-internet-gratuito-se-amplia-en-sectores-publicos-de-Ibarra-20230127-0002.html>
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press.
- Sadigzade, H. (2025). Can NotebookLM support English language learners? A theoretical perspective on AI tools in education. *Porta Universorum*, 1(6), 25–55.

- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.). Harvard University Press.

ANNEXES

APPENDICES A (Consent Form) & B(Pre-Study Survey) (Responses)

<https://docs.google.com/spreadsheets/d/1h9QHQyn57uXjxtanMYOhmsvFufKSxGvwE9NJVBmq6RU/edit?usp=sharing>

APPENDIX B - 114 AUTORIZACIÓN REALIZACIÓN DE INVESTIGACIÓN (ITCA UNIVERSITARIO)

https://drive.google.com/file/d/1AIU520_ZTYvCNaBsjszR9laepCHwdoPx/view?usp=drive_link

APPENDICES C - LIKERT SCALE (Control Group Post-Study Survey) (Responses)

<https://docs.google.com/spreadsheets/d/1G0e966rEhDRewNZFwqnRi0XmDc0LegEQ9u3v1-mO6mI/edit?usp=sharing>

APPENDICES D (Post-Study Survey) (Responses)

<https://docs.google.com/spreadsheets/d/1RE2SxIUnA7MkzIPNorevF0XekLWuFoptadSx-nxz9iM/edit?usp=sharing>

Appendix E - Field Observation & Conversation Analysis Protocol (FOCAP)

https://drive.google.com/file/d/1X1jwG_1yCLUC4uPk7KOnF_rHbjl4WoX3/view?usp=drivesdk