

Viabilidad pedagógica de la IA generativa de voz en la enseñanza del inglés*Pedagogical viability of generative AI voice in English language teaching***Angel González Escalante¹****Evelyn Guerrero Amarillas²**¹Autor de correspondencia: angelgonzalez@uas.edu.mx¹Universidad Autónoma de Sinaloa.
México.angelgonzalez@uas.edu.mx<https://orcid.org/0009-0007-6927-966X>²Universidad Autónoma de Sinaloa.
México.eve.guerrero.a@uas.edu.mx<https://orcid.org/0009-0009-6346-1211>**Datos del artículo:**

Recibido: junio 15, 2026

Aceptado: agosto 12, 2026

Publicación: septiembre 15, 2026

Palabras clave:bachillerato, inteligencia artificial,
innovación pedagógica, percepción,
pronunciación, síntesis de voz.**Keywords:**intelligence, pedagogical innovation,
perception, pronunciation, upper
secondary education, voice synthesis.DOI: <https://doi.org/10.53877/95a1xh30>

Este artículo está bajo la licencia

**Resumen**

La integración de la inteligencia artificial generativa en la producción de materiales educativos plantea interrogantes sobre su viabilidad pedagógica en contextos con restricciones institucionales. Este artículo documenta la experiencia de una institución pública de educación media superior que, ante la ausencia imprevista del componente auditivo en una serie de libros de inglés, recurrió a herramientas de inteligencia artificial generativa de voz para producir los materiales necesarios antes del inicio del ciclo escolar. El estudio adoptó un enfoque mixto con predominio cuantitativo, combinando la reconstrucción retrospectiva del proceso de producción con la aplicación de un cuestionario Likert a 47 docentes con experiencia previa en audios tradicionales. Los resultados muestran una percepción favorable en los cuatro bloques del instrumento ($M = 3.98$, $\alpha = 0.972$), con el valor pedagógico como dimensión mejor valorada y la coherencia curricular como aspecto de mayor consenso. El análisis cualitativo confirmó la utilidad del recurso para la práctica docente y desplazó las limitaciones señaladas hacia condiciones de infraestructura. Los hallazgos sugieren que la inteligencia artificial generativa de voz constituye una alternativa viable para la producción de materiales auditivos cuando se articula con criterios pedagógicos definidos, con implicaciones para instituciones con restricciones similares en América Latina.

Abstract

The integration of generative artificial intelligence in the production of educational materials raises questions about its pedagogical viability in contexts with institutional constraints. This article documents the experience of a public upper secondary education institution that, faced with the unexpected absence of an audio component in an English language textbook series, used generative AI voice tools to produce the required materials before the start of the school year. The study adopted a mixed-methods approach with quantitative predominance, combining a retrospective reconstruction of the production process with the administration of a Likert questionnaire to 47 teachers with prior experience using recorded audio materials. Results indicate a favorable and consistent perception across the four blocks of the instrument ($M = 3.98$, $\alpha = 0.972$), with pedagogical value as the highest-rated dimension and curricular coherence as the area of greatest consensus. Qualitative analysis of an open-ended question confirmed the resource's usefulness for teaching practice, while the limitations identified by teachers pointed to infrastructure conditions. Findings suggest that generative AI voice technology represents a viable alternative for the institutional production of audio materials when implemented

within a defined pedagogical framework, with relevant implications for institutions facing similar constraints across Latin America.

Forma sugerida de citar (APA):

González Escalante, A. y Guerrero Amarillas, E. (2026). Viabilidad pedagógica de la IA generativa de voz en la enseñanza del inglés. *RICEd: Revista de Investigación en Ciencias de la Educación*. 1(3). Ed. Esp. 24-37. <https://doi.org/10.53877/95a1xh30>

INTRODUCCIÓN

La inteligencia artificial generativa (IAG) refiere a sistemas computacionales capaces de producir contenido original como texto, imagen, audio o video, a partir del procesamiento de grandes volúmenes de datos y sin intervención humana directa en cada salida (García-Peñalvo et al., 2024). A diferencia de los sistemas de inteligencia artificial convencionales, que han sido diseñados para clasificar o predecir con base en patrones fijos, los sistemas generativos aprenden estructuras complejas del lenguaje y el sonido para producir resultados nuevos con alto grado de coherencia y naturalidad. En el ámbito educativo, su irrupción ha sido tan acelerada como desigual: mientras algunas instituciones la incorporan como herramienta pedagógica activa, muchas otras la enfrentan sin marcos claros de uso ni condiciones para hacerlo (Sánchez-Mendiola y Carbajal-Degante, 2023; Flores-Vivar y García-Peñalvo, 2023).

Ese panorama es especialmente visible en la producción de materiales didácticos. Las instituciones públicas de educación media superior en América Latina operan con restricciones presupuestarias, de tiempo y de personal especializado que dificultan el desarrollo de recursos de calidad (Sanabria et al., 2017; Sánchez-Mendiola et al., 2020). En el caso de los materiales de audio para la enseñanza del inglés, esas restricciones son particularmente críticas: contratar locutores profesionales, gestionar sesiones de grabación y garantizar la diversidad de voces, acentos y registros que exige una secuencia didáctica completa representa un costo que la mayoría de las instituciones públicas no puede asumir de forma sostenida. La consecuencia habitual es la ausencia de audio o el uso de materiales desactualizados que no responden a los programas vigentes (Miao y Holmes, 2023; Ramírez Martinell, 2022).

La IAG de voz ofrece una alternativa que hasta hace poco no existía a escala accesible. Los sistemas actuales de síntesis de voz neuronal permiten generar audio con múltiples perfiles vocales, variaciones de entonación, ritmo y acento, e incluso ambientación sonora, en fracciones del tiempo y el costo que implica una grabación convencional (Forero-Corba y Negre Bennasar, 2024; Rebolledo y Gisbert, 2025). Sin embargo, su adopción en la producción de materiales didácticos de inglés para el bachillerato es un fenómeno reciente y escasamente documentado. La literatura disponible sobre IAG en educación se concentra mayoritariamente en educación superior (Gallent-Torres et al., 2023; Martínez-Comesaña et al., 2023) y aborda con menor frecuencia aplicaciones específicas en la producción de audio para la enseñanza de lenguas extranjeras en niveles preuniversitarios (Briceño-Núñez et al., 2024). Ese vacío es precisamente el espacio que este trabajo busca documentar.

La relevancia de investigar este fenómeno se sostiene en tres frentes. Desde la perspectiva educativa, el componente auditivo es estructural en la enseñanza del inglés: sin exposición a input oral variado y contextualizado, el desarrollo de la comprensión auditiva queda comprometido desde el diseño del material. Desde la perspectiva tecnológica e institucional, la IAG representa una solución viable, rápida y escalable para instituciones que no pueden costear la producción convencional, siempre que los recursos generados cumplan

con criterios pedagógicos mínimos. Desde la perspectiva científica, la percepción docente sobre estos materiales constituye un indicador válido y ampliamente utilizado para evaluar la efectividad de recursos educativos en contextos reales (Cabero-Almenara y Palacios-Rodríguez, 2021), particularmente cuando el acceso directo al aprendizaje del estudiante no es el objeto central del estudio.

En ese marco, el presente artículo documenta una experiencia de innovación educativa ocurrida en una institución pública de educación media superior del noroeste de México, donde la ausencia imprevista del componente auditivo en materiales de inglés de reciente producción impulsó la adopción de IAG de voz como solución. El objetivo del estudio es describir la percepción de los docentes que utilizaron esos audios en su práctica, a partir de un cuestionario mixto aplicado a 47 profesores con experiencia previa en el uso de audios tradicionales grabados por humanos. Los hallazgos buscan aportar evidencia empírica sobre la viabilidad pedagógica de la IAG de voz en este contexto y ofrecer un referente replicable para instituciones que enfrenten condiciones similares.

Es pertinente destacar que la viabilidad pedagógica de ese recurso no puede evaluarse sin considerar el lugar que ocupa el audio en la enseñanza del inglés como lengua extranjera, la evolución tecnológica que hace posible su generación con inteligencia artificial y el papel que la percepción docente desempeña como indicador legítimo de efectividad en contextos donde el aprendizaje estudiantil no constituye el objeto central del estudio.

1.1 El audio como componente estructural en la enseñanza del inglés

La comprensión auditiva no es un complemento opcional en la enseñanza del inglés como lengua extranjera: es una habilidad receptiva activa que condiciona el desarrollo de las demás competencias comunicativas. Escuchar no equivale a oír; implica procesar información fonológica, léxica y pragmática en tiempo real, distinguir acentos, interpretar pausas y anticipar significados a partir del contexto (Córdoba Cubillo et al., 2005). Para que ese proceso ocurra en el aula, los materiales deben proveer input oral suficiente, variado y contextualizado.

Martínez Molina et al. (2020) documentan que la exposición sistemática a audio en lengua extranjera, cuando está bien articulada con actividades de comprensión, mejora no solo la habilidad auditiva sino también la pronunciación y la confianza del estudiante para interactuar en el idioma. Esta evidencia refuerza la idea de que el audio no puede pensarse como un recurso accesorio: su ausencia o su baja calidad compromete la secuencia didáctica desde su diseño. En ese sentido, la revisión de Rebolledo y Gisbert (2025) sobre herramientas de inteligencia artificial aplicadas al aprendizaje del inglés confirma que la calidad y variedad del input auditivo sigue siendo una variable central en el rendimiento de los estudiantes de lengua extranjera, independientemente del medio tecnológico que lo produzca.

1.2 De la síntesis de voz tradicional a la inteligencia artificial generativa de voz

Los sistemas de texto a voz (TTS, por sus siglas en inglés) existen desde hace décadas, pero su evolución no ha sido lineal. Los modelos tradicionales basados en concatenación de segmentos de voz humana producían locuciones reconocibles pero mecánicas, con entonación plana y escasa variación expresiva. Esa limitación los hizo poco atractivos para usos pedagógicos donde la naturalidad del habla es una variable relevante (Bione y Cardoso, 2020).

La aparición de modelos neuronales de síntesis de voz transformó ese panorama. Los sistemas actuales de IAG de voz no concatenan fragmentos pregrabados sino que generan formas de onda completas a partir de representaciones aprendidas del habla humana, lo que

les permite reproducir variaciones de ritmo, entonación, acento y registro con un nivel de naturalidad que los modelos anteriores no podían alcanzar (Nussbaum et al., 2025). García-Peñalvo et al. (2024) sitúan este avance dentro de la transformación más amplia que la IAG está produciendo en la educación, señalando que su capacidad generativa abre posibilidades que van más allá de la automatización de tareas y alcanzan la producción de recursos pedagógicos de calidad. Forero-Corba y Negre Bennasar (2024) confirman esa tendencia en su revisión sistemática sobre aplicaciones de inteligencia artificial en educación, donde la síntesis de voz neuronal aparece como una de las herramientas con mayor proyección para el diseño de materiales.

Desde la perspectiva del aprendizaje del inglés, Tai y Chen (2024) demostraron en un ensayo controlado que el uso de voz generada por inteligencia artificial en actividades de comprensión auditiva no solo no perjudica el rendimiento de los estudiantes, sino que puede reducir la ansiedad asociada al *listening*, particularmente cuando las voces presentan variedad de perfiles y ritmos. Ese hallazgo es relevante para el presente estudio porque sugiere que la naturalidad percibida por los docentes tiene correlato en la experiencia del estudiante.

1.3 Innovación educativa situada: restricciones como motor de cambio

La innovación educativa no ocurre en el vacío. Martínez-Bonafé y Rogero-Anaya (2021) argumentan que las condiciones del entorno institucional, incluyendo sus restricciones presupuestarias, sus urgencias operativas y sus marcos normativos, son factores que no solo condicionan, sino que en muchos casos detonan procesos de cambio que no habrían surgido en condiciones ideales. Esa perspectiva sitúa la innovación como un fenómeno emergente, inseparable del contexto en que ocurre.

En el caso de México, el rediseño curricular impulsado por la Nueva Escuela Mexicana representó una presión institucional de gran alcance que obligó a numerosas instancias educativas a replantear sus materiales y prácticas en plazos muy ajustados (Ramírez Martinell, 2022). Área-Moreira y Adell-Segura (2021) advierten que en ese tipo de procesos la tecnología digital no es neutral: su adopción responde a condiciones concretas de posibilidad y no garantiza por sí sola la mejora educativa. Lo que determina el valor pedagógico de una herramienta tecnológica es la coherencia entre sus capacidades y los objetivos del proceso en que se inserta. Domingo-Coscollola et al. (2020) refuerzan esta idea al documentar que los docentes que protagonizan innovaciones desde dentro de sus instituciones tienden a valorar más la funcionalidad práctica del recurso que su sofisticación técnica, lo cual tiene implicaciones directas para interpretar los resultados de este estudio.

1.4 La percepción docente como indicador de efectividad de materiales

Evaluar la efectividad de un material didáctico sin medir directamente el aprendizaje del estudiante es una práctica metodológica legítima cuando el objeto de estudio es la experiencia del docente como usuario del recurso. Cepeda-Romero et al. (2017) establecen que la evaluación de materiales didácticos digitales desde la perspectiva del profesorado permite identificar dimensiones de calidad que las pruebas de rendimiento estudiantil no capturan, como la integración curricular, la facilidad de uso o la coherencia con la práctica de aula.

Cabero-Almenara y Palacios-Rodríguez (2021) documentan que la valoración docente de recursos tecnológicos educativos es un indicador ampliamente validado en la literatura, utilizado de forma recurrente en investigaciones sobre adopción de tecnología en contextos de enseñanza formal. Aguilar-Martínez et al. (2024) confirman esa validez en un estudio reciente sobre percepciones docentes hacia tecnología educativa, demostrando que los

cuestionarios Likert aplicados a profesores producen datos consistentes y metodológicamente defendibles cuando el instrumento está bien construido y la muestra tiene experiencia directa con el objeto evaluado. Navarro-Carrascosa (2024) añade una dimensión específica al señalar que en la enseñanza de lenguas la percepción docente sobre los materiales de audio es especialmente relevante porque el profesor es quien media entre el recurso y el estudiante, decidiendo cómo, cuándo y con qué propósito se usa. Wang et al. (2024) respaldan esta postura desde una revisión sistemática amplia sobre inteligencia artificial en educación, donde la percepción de los actores del proceso emerge como variable central para comprender la adopción real de las herramientas.

1.5 Colaboración interdisciplinaria en la producción de materiales educativos

La producción de materiales didácticos de calidad rara vez es obra de un solo perfil profesional. Prendes-Espinosa y Cerdán-Cartagena (2021) argumentan que los proyectos de innovación educativa que involucran actores con competencias técnicas diversas tienden a producir resultados de mayor calidad y sostenibilidad que los desarrollados desde un único campo de *expertise*. Esa diversidad no requiere necesariamente de estructuras formales: Ayuso-del Puerto y Gutiérrez-Esteban (2022) documentan que incluso actores en formación pueden protagonizar procesos de innovación con herramientas de inteligencia artificial cuando se les da una tarea concreta con consecuencias reales.

Barrios Alvarado (2025) ilustra este principio en un caso de colaboración entre diseñadores y docentes para la producción de material didáctico en educación secundaria, concluyendo que la convergencia de saberes técnicos y pedagógicos eleva la calidad del producto final y genera aprendizaje en todos los participantes. Esa dinámica es reconocible en la experiencia documentada en este artículo, donde la incorporación de un estudiante universitario con habilidades en producción multimedia al proceso de generación de audios produjo una mejora cualitativa que el equipo académico no habría podido lograr por sus propios medios.

Cabe señalar que el diseño de estudio de caso instrumental empleado en esta investigación sigue los principios establecidos por Stake (1999), referente canónico e histórico en la investigación educativa cualitativa, cuya tipología distingue entre casos intrínsecos, instrumentales y colectivos según la relación entre el caso y el fenómeno que se busca comprender. El presente estudio es instrumental en tanto el caso específico de una institución pública de educación media superior sirve para iluminar un problema de mayor alcance: la viabilidad de la IAG de voz como herramienta institucional para la producción de materiales de audio en el bachillerato.

MÉTODOS Y MATERIALES

Diseño de la investigación

Este estudio adoptó un enfoque metodológico mixto con predominio cuantitativo (Pereira-Pérez, 2011), articulado en dos movimientos complementarios. El primero consistió en la reconstrucción descriptiva del proceso de producción de los materiales auditivos, a partir de fuentes documentales, registros digitales y narrativas del equipo participante. Dado que dicho proceso precedió al diseño formal de la investigación, su descripción se construyó retrospectivamente mediante triangulación de fuentes (Forni y De Grande, 2020), sin manipular ni reinterpretar los hechos ocurridos. El segundo movimiento se centró en la

evaluación del producto resultante a través de un cuestionario aplicado a los docentes que lo utilizaron en condiciones reales de aula.

El diseño global corresponde a un estudio de caso instrumental (Stake, 1999), en tanto el caso particular opera como vehículo para comprender un fenómeno de mayor alcance: la viabilidad pedagógica e institucional de la inteligencia artificial generativa de voz como solución escalable en la producción de materiales didácticos de inglés para el bachillerato. El diseño fue no experimental y transversal, con recopilación de datos en un único momento sin intervención sobre las variables estudiadas.

Contexto y participantes

La investigación se desarrolló en una institución pública de educación media superior del noroeste de México, en el marco del rediseño curricular derivado de la implementación de la Nueva Escuela Mexicana. Ese rediseño exigió la producción de nuevos materiales didácticos para la asignatura de inglés del primer semestre, lo que desencadenó el proceso de innovación que este artículo documenta y evalúa.

Los participantes fueron 47 docentes que impartían la asignatura de inglés del primer semestre con los materiales producidos por la institución al momento de la recolección de datos. El criterio de inclusión fue doble: haber utilizado los audios generados con inteligencia artificial generativa durante al menos un semestre completo y contar con experiencia previa en el uso de audios tradicionales grabados por humanos. Este segundo criterio se estableció a propósito para garantizar que los participantes dispusieran de un referente comparativo real, lo que otorga mayor solidez interpretativa a sus valoraciones. La invitación se distribuyó a través del canal institucional de la academia de inglés, con participación voluntaria y consentimiento informado digital.

Proceso de producción de los audios

A pocos días del inicio del ciclo escolar en que los materiales serían utilizados por primera vez, el equipo académico enfrentó un escenario imprevisto: los audios, ya contemplados en el diseño didáctico, no serían producidos por los medios originalmente acordados debido a decisiones externas. Esto planteó la necesidad de resolver el componente auditivo en un margen de tiempo muy ajustado, descartando de inmediato las vías convencionales por restricciones simultáneas de tiempo y presupuesto.

La respuesta se articuló en dos fases. En la primera, el equipo generó los audios con una herramienta de inteligencia artificial generativa de voz mediante ajustes básicos de configuración, principalmente la selección del género de la voz. Ese material cubrió la necesidad inmediata al inicio de las clases. En la segunda fase, un estudiante universitario de la misma institución con habilidades en producción multimedia trabajó una versión mejorada utilizando una herramienta especializada distinta. Sobre esa versión incorporó variaciones de ritmo y entonación, diversidad de perfiles de voz y ambientación sonora contextualizada en los diálogos, logrando un nivel notablemente mayor de naturalidad. Los audios resultantes de esta segunda fase sustituyeron gradualmente a los de la primera durante las semanas iniciales del semestre.

Por decisión metodológica, los nombres de las herramientas utilizadas en ambas fases no se especifican en el cuerpo del artículo. Su relevancia científica radica en las capacidades funcionales que ofrecen, síntesis de voz neuronal, control de ritmo y entonación y diversidad de perfiles de voz, y no en marcas comerciales sujetas a cambios de disponibilidad o modelo de negocio.

Instrumento

Se diseñó un cuestionario ad hoc compuesto por 15 ítems en escala Likert de cinco puntos (1 = Totalmente en desacuerdo, 5 = Totalmente de acuerdo) y una pregunta abierta final. Los ítems se organizaron en cuatro bloques temáticos: el Bloque A sobre percepción general de los audios con inteligencia artificial (ítems 1 a 4); el Bloque B sobre valor pedagógico y utilidad en el aula (ítems 5 a 8); el Bloque C sobre experiencia y adecuación al contexto (ítems 9 a 12); y el Bloque D sobre implementación institucional y percepción futura (ítems 13 a 15). La pregunta abierta (ítem 16) recogió narrativas sobre la experiencia general de uso en la práctica docente y se analizó de forma independiente.

Antes de la aplicación definitiva, el instrumento fue sometido a una prueba piloto con un grupo reducido de docentes para verificar la comprensión de los ítems y el tiempo promedio de respuesta. Los ajustes derivados de esa prueba se incorporaron a la versión final. La consistencia interna del instrumento completo se calculó mediante el coeficiente alfa de Cronbach, obteniendo $\alpha = 0.972$, lo que indica alta homogeneidad entre los ítems. Los coeficientes por bloque oscilaron entre 0.870 y 0.955, confirmando consistencia adecuada en cada dimensión.

Análisis de datos

El componente cuantitativo se analizó mediante estadística descriptiva: frecuencias absolutas y porcentajes por ítem, medias y desviaciones estándar por ítem y por bloque, y media global del instrumento. Los resultados se presentan en tablas comparativas que facilitan la identificación de patrones de respuesta entre bloques.

El componente cualitativo del ítem 16 se procesó mediante análisis de contenido temático (Viramontes Anaya, 2024), identificando categorías emergentes en las respuestas de los docentes, calculando su frecuencia de aparición y contrastándolas con los resultados cuantitativos. Esa triangulación permitió establecer convergencias, matices y divergencias entre ambas fuentes de datos, construyendo una lectura más completa de las percepciones docentes que la que ofrecen los datos numéricos de forma aislada.

RESULTADOS

Resultados cuantitativos

El cuestionario fue respondido por 47 docentes. La media global del instrumento fue de 3.98 (DE = 0.87), con un coeficiente alfa de Cronbach de $\alpha = 0.972$, lo que indica alta consistencia interna. Los coeficientes por bloque oscilaron entre 0.870 y 0.955, confirmando homogeneidad adecuada en cada dimensión. La Tabla 1 presenta las medias y desviaciones estándar por ítem.

Tabla 1
Medias y desviaciones estándar por ítem (n = 47)

Ítem	Descripción	M	DE
1	Naturalidad imperceptible como IA	3.81	1.04
2	Pronunciación adecuada para fines didácticos	4.06	0.89
3	Calidad de sonido adecuada para el aula	3.98	0.92
4	Ritmo y entonación favorecen la comprensión	3.94	0.89
5	Permiten trabajar comprensión auditiva efectivamente	4.13	0.71
6	Facilitan exposición al inglés auténtico	4.04	0.86
7	Complementan bien el resto de los materiales	4.04	0.84

8	Su disponibilidad mejora la planeación docente	4.09	0.78
9	Se integran coherentemente con el libro de texto	4.11	0.60
10	Generan interés o curiosidad en los estudiantes	3.83	0.92
11	Resultan más prácticos que materiales anteriores	3.85	0.93
12	Facilitan la secuencia pedagógica de las clases	3.98	0.85
13	Han favorecido continuidad y consistencia en clases	3.91	0.86
14	Funcionan adecuadamente con distintos grupos	3.85	0.93
15	Es positivo que la institución promueva el uso de IA	4.09	0.93

Nota: Escala de 1 (Totalmente en desacuerdo) a 5 (Totalmente de acuerdo).

Los datos por ítem revelan un rango de medias relativamente estrecho, de 3.81 a 4.13, lo que sugiere una valoración positiva sostenida sin picos extremos en ninguna dirección. Para facilitar la lectura de estos patrones a nivel dimensional, la Tabla 2 presenta los estadísticos agregados por bloque junto con los coeficientes de consistencia interna correspondientes.

Tabla 2

Estadísticos por bloque (n = 47)

Bloque	Contenido	M	DE	α
A	Percepción general de los audios con IA	3.95	0.93	0.913
B	Valor pedagógico y utilidad en el aula	4.07	0.79	0.955
C	Experiencia y adecuación al contexto	3.94	0.83	0.870
D	Implementación institucional y percepción futura	3.95	0.90	0.906

Nota: α = coeficiente alfa de Cronbach por bloque.

Los resultados muestran una tendencia positiva consistente en los cuatro bloques, con medias que oscilan entre 3.94 y 4.07. El Bloque B concentra las valoraciones más altas del instrumento, con el ítem 5 alcanzando la media más elevada ($M = 4.13$, $DE = 0.71$), lo que indica alta concentración de respuestas en las categorías de acuerdo. El ítem 9 presenta la menor dispersión de todo el instrumento ($DE = 0.60$), reflejando el mayor nivel de consenso entre los docentes respecto a la coherencia entre los audios y las actividades del libro de texto.

En el extremo opuesto, los ítems 1, 10 y 11 registran las medias más bajas del instrumento, con valores de 3.81, 3.83 y 3.85 respectivamente. El ítem 1 es además el único con una desviación estándar superior a 1.0 ($DE = 1.04$), lo que indica mayor heterogeneidad en las respuestas sobre la capacidad de los audios de pasar como voz humana. Los ítems 10 y 11 muestran menor acuerdo relativo en aspectos relacionados con el interés estudiantil y la comparación con materiales anteriores.

Resultados cualitativos

El análisis de contenido temático de las respuestas al ítem 16 permitió identificar cinco categorías emergentes. La Tabla 3 presenta su distribución por frecuencia.

Tabla 3

Categorías temáticas del ítem 16 (n = 47)

Categoría	Frecuencia	%
1. Utilidad pedagógica y apoyo a la práctica docente	24	51.1
2. Apoyo al desarrollo de la comprensión auditiva	6	12.8
3. Limitaciones técnicas y condiciones de uso	6	12.8
4. Naturalidad y pronunciación percibida	5	10.6
5. Valoración positiva con sugerencias de mejora	6	12.8

Nota: Una respuesta puede contener más de una categoría.

La categoría de mayor frecuencia fue la utilidad pedagógica (51.1%), cuyos comentarios describen los audios como un recurso que apoya la explicación docente y dinamiza las clases. Esta tendencia converge con los resultados del Bloque B, particularmente con el ítem 5, que registró la media más alta del instrumento.

Las categorías 2 y 3 presentaron la misma frecuencia (12.8%). La categoría 2 refuerza los resultados del Bloque B al destacar el apoyo de los audios al desarrollo de la comprensión auditiva y la exposición al inglés. La categoría 3, en cambio, introduce un matiz ausente en los datos cuantitativos: las observaciones sobre limitaciones técnicas apuntan a condiciones de implementación como el volumen bajo o la acústica de las aulas, no a deficiencias del recurso en sí mismo.

La categoría 4 (10.6%) representa la divergencia más relevante respecto a los datos cuantitativos. Algunos docentes describieron ciertas voces como robotizadas o poco naturales, perspectiva que contrasta con las valoraciones mayoritariamente positivas del Bloque A pero que es consistente con la mayor dispersión observada en el ítem 1 ($DE = 1.04$). Entre las respuestas de esta categoría destaca un caso aislado en que el docente señaló haber dejado de utilizar el recurso por considerarlo artificial. Aunque minoritaria, esta respuesta es relevante porque evidencia que la heterogeneidad estadística del ítem 1 tiene correlato cualitativo real. La categoría 5 (12.8%) refleja una actitud crítica constructiva: los docentes reconocen el valor del recurso y señalan áreas de mejora específicas, lo que sugiere apropiación del material y disposición hacia su perfeccionamiento.

DISCUSIÓN

Los resultados del estudio muestran una percepción docente favorable y consistente hacia los audios generados con inteligencia artificial generativa de voz, con una media global de 3.98 y coeficientes de consistencia interna que confirman la solidez del instrumento. Esa valoración positiva no se concentra en una sola dimensión, sino que se distribuye de forma homogénea entre los cuatro bloques, con un rango de medias agregadas de apenas 0.13 puntos. Este patrón es coherente con lo documentado por Wang et al. (2024) en su revisión sistemática sobre inteligencia artificial en educación, donde la aceptación de herramientas de IA por parte de los docentes tiende a ser más alta cuando el recurso resuelve una necesidad pedagógica concreta y se integra de forma natural en la práctica existente.

El Bloque B, referido al valor pedagógico y la utilidad en el aula, concentró las valoraciones más altas del instrumento. El ítem 5, sobre la capacidad de los audios para trabajar la comprensión auditiva de forma efectiva, alcanzó la media más elevada de todo el cuestionario ($M = 4.13$). Este resultado es consistente con los hallazgos de Tai y Chen (2024), quienes demostraron que el audio generado por inteligencia artificial no perjudica el rendimiento en tareas de comprensión auditiva en inglés como lengua extranjera y puede incluso reducir la ansiedad asociada a ese tipo de actividades. La preeminencia del valor funcional sobre otros aspectos del recurso sugiere que los docentes evaluaron los audios principalmente desde su utilidad práctica en el aula, más que desde criterios de perfección técnica o estética vocal. Esa distinción entre utilidad funcional y perfección estética es relevante para interpretar los resultados en su conjunto, porque explica por qué valoraciones moderadas en algunos ítems coexisten con una aprobación general sólida.

El ítem 9, sobre la coherencia entre los audios y las actividades del libro de texto, presentó la menor dispersión de todo el instrumento ($DE = 0.60$), lo que indica el mayor nivel de consenso entre los participantes. Ese dato tiene implicaciones curriculares directas: sugiere que el proceso de producción logró articular el componente auditivo con el resto del material de forma que los docentes perciben como coherente. Cepeda-Romero et al. (2017) señalan que

la integración curricular es una de las dimensiones más valoradas por el profesorado al evaluar materiales didácticos digitales, precisamente porque afecta de forma directa la planificación y ejecución de las clases. Que sea el ítem de mayor consenso en este estudio refuerza la idea de que la coherencia entre el audio y el material escrito es un criterio prioritario para los docentes, por encima incluso de la naturalidad de las voces.

En el extremo opuesto, el ítem 1 registró la media más baja del instrumento ($M = 3.81$) y la única desviación estándar superior a 1.0 ($DE = 1.04$), lo que indica heterogeneidad real en la percepción sobre la capacidad de los audios de pasar como voz humana. Este resultado es esperado a la luz de la literatura sobre síntesis de voz neuronal: Nussbaum et al. (2025) documentan que la naturalidad percibida de una voz sintética es una variable subjetiva altamente sensible a la experiencia previa del oyente, su familiaridad con voces artificiales y el contexto en que se produce la escucha. Bione y Cardoso (2020) añaden que los docentes de lenguas extranjeras tienden a ser más críticos que los estudiantes respecto a la naturalidad de las voces sintéticas, precisamente porque su formación los sensibiliza hacia matices fonéticos que un oyente no especializado no detecta. La dispersión del ítem 1 refleja esa variabilidad individual y no debe interpretarse como rechazo al recurso, sino como un diagnóstico técnico sobre un aspecto mejorable. El análisis cualitativo del ítem 16 confirma esta lectura: las respuestas de la categoría 4, referidas a la naturalidad percibida, describen reservas específicas sobre ciertas voces sin que ninguna de ellas suponga un cuestionamiento del recurso en su conjunto, con la excepción de un caso aislado.

Los ítems 10 y 11, sobre el interés estudiantil y la comparación con materiales anteriores, presentaron las medias más bajas junto con el ítem 1. Una interpretación posible es que los docentes son más cautelosos cuando sus valoraciones implican hablar en nombre de sus estudiantes o establecer comparaciones directas con recursos previos, lo que introduce un sesgo de prudencia que no necesariamente refleja una valoración negativa del material. Aguilar-Martínez et al. (2024) documentan un patrón similar en estudios sobre percepción docente hacia tecnología educativa, donde las dimensiones que requieren inferir la experiencia del estudiante producen sistemáticamente respuestas más conservadoras que las dimensiones centradas en la experiencia del propio docente.

El análisis cualitativo aporta una dimensión que los datos numéricos no capturan completamente. La categoría 3, referida a limitaciones técnicas y condiciones de uso, desplaza la crítica del recurso hacia el contexto de implementación: volumen insuficiente, acústica deficiente del aula o dificultades de acceso desde el celular. Fernández-Gutiérrez et al. (2020) argumentan que la efectividad percibida de las tecnologías educativas depende tanto de las características del recurso como de las condiciones institucionales en que se usa, y que separar ambas variables es indispensable para una evaluación justa. En este caso, esa distinción es favorable para el recurso: las limitaciones señaladas por los docentes son de infraestructura, no de calidad del audio.

Desde una perspectiva institucional, la experiencia documentada ilustra lo que Martínez-Bonafé y Rogero-Anaya (2021) describen como innovación educativa emergente: un proceso que no surgió de una planificación deliberada sino de una restricción contextual que el equipo convirtió en oportunidad. La incorporación de un estudiante universitario al proceso de producción es coherente con lo planteado por Prendes-Espinosa y Cerdán-Cartagena (2021), quienes señalan que los proyectos de innovación que integran actores con perfiles técnicos diversos tienden a producir resultados de mayor calidad y sostenibilidad. Que ese actor fuera un estudiante en formación, sin experiencia profesional formal en el área, añade un elemento que merece atención en futuras investigaciones sobre colaboración interdisciplinaria en la producción de materiales educativos.

Finalmente, los resultados de este estudio se suman a una evidencia creciente sobre la viabilidad de la IAG como herramienta institucional para la producción de recursos educativos en contextos con restricciones de tiempo y presupuesto (Miao y Holmes, 2023; García-Peñalvo et al., 2024). La percepción docente favorable, sostenida en los cuatro bloques del instrumento y respaldada por los datos cualitativos, sugiere que cuando la IAG de voz se articula con criterios pedagógicos claros y se integra coherentemente con el material existente, constituye una alternativa viable y escalable a los modelos convencionales de producción de audio didáctico.

CONCLUSIONES

Este estudio documenta una experiencia de innovación educativa en la que la inteligencia artificial generativa de voz resolvió de forma viable, rápida y pedagógicamente aceptable la ausencia imprevista del componente auditivo en materiales de inglés para el bachillerato. La percepción de 47 docentes con experiencia previa en el uso de audios tradicionales grabados por humanos fue favorable en los cuatro bloques del instrumento, con una media global de 3.98 y alta consistencia interna. Esa valoración no se concentró en una sola dimensión, sino que se distribuyó de forma homogénea, lo que sugiere una aprobación integral del recurso y no solo de alguno de sus aspectos particulares.

La aportación central de este trabajo es de carácter práctico e institucional. Ante condiciones que combinan urgencia temporal, restricciones presupuestarias y necesidad de escalabilidad, la IAG de voz demostró ser una alternativa funcional que los docentes reconocen como útil para su práctica. El proceso se escaló exitosamente a los tres tomos subsecuentes de la serie, lo que confirma la sostenibilidad del modelo más allá de la solución inicial. Ese trayecto, de la contingencia a la consolidación, constituye un referente replicable para instituciones públicas de educación media superior que enfrenten condiciones similares en la producción de materiales auditivos para la enseñanza del inglés.

Desde la perspectiva científica, el estudio aporta evidencia empírica en un área con escasa documentación en la literatura hispanohablante: el uso de IAG de voz específicamente para la producción de materiales didácticos de audio en el bachillerato latinoamericano. Los resultados contribuyen a la discusión sobre los criterios con que los docentes evalúan recursos tecnológicos en contextos reales, confirmando que la utilidad funcional y la coherencia curricular pesan más que la perfección técnica en la valoración cotidiana del material.

Limitaciones

El estudio presenta limitaciones que condicionan el alcance de sus conclusiones. El diseño transversal recoge percepciones en un único momento y no permite observar cambios en la práctica docente ni efectos sostenidos en el tiempo. La muestra, aunque pertinente y bien definida, corresponde a una sola institución, lo que limita la generalización de los resultados a otros contextos sin la mediación de estudios comparativos. Por último, el instrumento mide la percepción docente pero no el aprendizaje de los estudiantes, de modo que la relación entre el uso de los audios y la mejora en la comprensión auditiva no está establecida en este trabajo y constituye una pregunta abierta para investigaciones futuras.

Líneas futuras de investigación

Los hallazgos de este estudio abren cuatro líneas de investigación que merecen atención. La primera es el diseño de estudios longitudinales que permitan observar el efecto sostenido del

uso de audios con IAG en el desempeño de los estudiantes en comprensión auditiva y producción oral. La segunda es la comparación sistemática entre audios generados con IAG y grabaciones de hablantes humanos en condiciones controladas, para establecer con mayor precisión en qué dimensiones y para qué perfiles de estudiante la diferencia es pedagógicamente relevante. La tercera es la incorporación de la voz del estudiante al análisis, explorando su percepción y motivación ante este tipo de materiales. La cuarta es la evaluación de la transferibilidad del modelo a otros niveles educativos y otras asignaturas, analizando su costo-beneficio institucional en contextos de producción de materiales a escala.

Más allá de la innovación tecnológica, la relevancia de la IA generativa en la producción de materiales didácticos reside en su articulación con el criterio del profesorado. Los hallazgos respaldan que esta tecnología es una alternativa pedagógicamente sólida para contextos con limitaciones institucionales, siempre que el docente mantenga el rol central en la curaduría y validación de los recursos auditivos resultantes. La evidencia presentada no agota la discusión sobre el tema, pero contribuye a fundamentarla con datos empíricos provenientes de un contexto latinoamericano escasamente documentado en la literatura especializada.

REFERENCIAS BIBLIOGRÁFICAS

- Abreus González, P. A., Martínez Molina, P. D., & Castiñeira Enríquez, P. A. (2020). El desarrollo de la comprensión auditiva en inglés mediante la audición extensiva. *Qualitas Revista Científica*, 19(19), 077-094. <https://bit.ly/3P8w2fp>
- Aguiar-Martínez, J. L., Sánchez-Prieto, J. C., & Martínez-Abad, F. (2024). Percepciones docentes hacia los efectos de la brecha digital y la inclusión educativa. *Revista Latinoamericana de Tecnología Educativa – RELATEC*, 23(2), 51-67. <https://acortar.link/HSGueU>
- Área-Moreira, M., & Adell-Segura, J. (2021). Tecnologías digitales y cambio educativo. Una aproximación crítica. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 19(4), 83-96. <https://acortar.link/xJTwiF>
- Ayuso-del Puerto, D., & Gutiérrez-Esteban, P. (2022). La inteligencia artificial como recurso educativo durante la formación inicial del profesorado. *RIED. Revista Iberoamericana de Educación a Distancia*, 25(2), 347-362. <https://acortar.link/MVH7EV>
- Barrios Alvarado, A. G. (2025). Práctica interdisciplinaria entre el diseño y la educación para el diseño de material didáctico destinado a alumnos de educación secundaria. *Revista de Estudios Interdisciplinarios del Arte, Diseño y la Cultura*, (15), 347-371. <http://bit.ly/4dcb2hK>
- Bione, T., & Cardoso, W. (2020). Synthetic voices in the foreign language context. *Language Learning & Technology*, 24(1), 169-186. <https://acortar.link/OIV1Uq>
- Briceño-Núñez, C. E., Pernia-Molina, R., Teixeira-Robles, Y. M., & Juárez-Viloria, C. N. (2024). Enfoque holístico en la enseñanza de idiomas: usos de la inteligencia artificial en contextos universitarios. *DIDAC*, (84 JUL-DIC), 30-43. <https://acortar.link/56E9iy>. 84_JUL-DIC.210
- Cabero-Almenara, J., & Palacios-Rodríguez, A. (2021). La evaluación de la educación virtual: las e-actividades. *RIED. Revista Iberoamericana de Educación a Distancia*, 24(2), 169-188. <https://acortar.link/EdwChD>
- Cepeda-Romero, O., Gallardo-Fernández, I. M., & Rodríguez-Rodríguez, J. (2017). La evaluación de los materiales didácticos digitales. *Revista Latinoamericana de Tecnología Educativa – RELATEC*, 16(2), 79-95. <https://acortar.link/PrWMns>

- Córdoba Cubillo, P., Coto Keith, R., & Ramírez Salas, M. (2005). La comprensión auditiva: definición, importancia, características, procesos, materiales y actividades. *Actualidades Investigativas en Educación*, 5(1). <https://acortar.link/F2QBqk>
- Fernández-Gutiérrez, M., Gimenez, G., & Calero, J. (2020). Is the use of ICT in education leading to higher student outcomes? Analysis from the Spanish autonomous communities. *Computers & Education*, 157, 103969. <https://acortar.link/zr6DQB>
- Flores-Vivar, J. M., & García-Peñalvo, F. J. (2023). Reflexiones sobre la ética, potencialidades y retos de la inteligencia artificial en el marco de la educación de calidad (ODS4). *Comunicar*, 31(74), 37–47. <https://acortar.link/Y8QBtz>
- Forero-Corba, W., & Negre Bennasar, F. (2024). Técnicas y aplicaciones del machine learning e inteligencia artificial en educación: una revisión sistemática. *RIED. Revista Iberoamericana de Educación a Distancia*, 27(1), 209–253. <https://acortar.link/W0ituF>
- Forni, P., & De Grande, P. (2020). Triangulación y métodos mixtos en las ciencias sociales contemporáneas. *Revista Mexicana de Sociología*, 82(1), 159–189. <https://acortar.link/ucBAKp>
- Gallent-Torres, C., Zapata-González, A., & Ortego-Hernando, J. L. (2023). El impacto de la inteligencia artificial generativa en educación superior: una mirada desde la ética y la integridad académica. *RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa*, 29(2), art. M5. <https://acortar.link/8P5FcV>
- García-Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED. Revista Iberoamericana de Educación a Distancia*, 27(1), 9–39. <https://acortar.link/l26tXd>
- Martínez-Bonafé, J., & Rogero-Anaya, J. (2021). El entorno y la innovación educativa. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 19(4), 71–81. <https://acortar.link/lpeONH>
- Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocaranza-Prado, I., & Kreibel, D. (2023). Impacto de la inteligencia artificial en los métodos de evaluación en educación primaria y secundaria: revisión sistemática de la literatura. *Revista de Psicodidáctica*, 28(2), 93–103. <https://acortar.link/SoCDwN>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://acortar.link/5GOvy9>
- Nussbaum, C., Frühholz, S., & Schweinberger, S. R. (2025). Understanding voice naturalness. *Trends in Cognitive Sciences*, 29(5), 467–480. <https://acortar.link/5zbkOn>
- Pereira-Pérez, Z. (2011). Los diseños de método mixto en la investigación en educación: una experiencia concreta. *Revista Electrónica Educare*, 15(1), 15–29. <https://acortar.link/nu9j8j>
- Prendes-Espinosa, M. P., & Cerdán-Cartagena, F. (2021). Tecnologías avanzadas para afrontar el reto de la innovación educativa. *RIED. Revista Iberoamericana de Educación a Distancia*, 24(1), 35–53. <https://acortar.link/jWBSOD>
- Ramírez Martinell, A. (2022). Cultura digital y educación media superior. En A. Aguirre y A. Ramírez (Coords.), *La educación media superior ante la nueva escuela mexicana* (pp. 66–79). COBAEV. <https://bit.ly/4spGJsJ>
- Rebolledo, R., & Gisbert, M. (2025). Aprendizaje adaptativo del inglés como lengua extranjera con herramientas de inteligencia artificial: una revisión sistemática de la literatura. *Profesorado. Revista de Currículum y Formación del Profesorado*, 29(1), 241–264. <https://acortar.link/Espebo>
- Sanabria, A., Álvarez-Núñez, Q., & Peirats, J. (2017). Las políticas educativas en la producción y distribución de materiales didácticos digitales. *Revista Latinoamericana de Tecnología Educativa – RELATEC*, 16(2), 63–77. <https://acortar.link/OKJR0v>

- Sánchez-Mendiola, M., & Carbajal-Degante, E. (2023). La inteligencia artificial generativa y la educación universitaria: ¿salió el genio de la lámpara? *Perfiles Educativos*, 45(Especial), 70–86. <https://acortar.link/pwbSKr>
- Sánchez-Mendiola, M., Martínez-Franco, A. I., Torres-Carrasco, U., de Agüero-Servín, M., Hernández-Romo, A. K., Benavides-Lara, M. A., Rendón-Cazales, V. J., & Jaimes-Vergara, C. A. (2020). Retos educativos durante la pandemia de COVID-19: una encuesta a profesores de la UNAM. *Revista Digital Universitaria*, 21(3). <https://acortar.link/PnKsdF>
- Stake, R. E. (1999). *Investigación con estudio de casos* (R. Filella, Trad., 2.^a ed.). Ediciones Morata.
- Tai, T.-Y., & Chen, H. H.-J. (2024). The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: a randomized controlled trial. *Humanities and Social Sciences Communications*, 11, art. 447. <https://acortar.link/fBYIHG>
- Viramontes Anaya, E. (2024). Análisis cualitativo en la investigación. *IE Revista de Investigación Educativa de la REDIECH*, 15, e2074. <https://acortar.link/uBm5ju>
- Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252 (Part A), 124167. <https://acortar.link/xddfuO>