

Estrategias didácticas innovadoras y producción oral en inglés como lengua extranjera: revisión de alcance con búsqueda sistematizada (2019–2025)

Innovative Teaching Strategies and Oral Production in English as a Foreign Language: A Scoping Review with a Systematized Search (2019–2025)

Carmen Anelí Goicochea Fabián¹ 

¹ Universidad Nacional de Trujillo, Trujillo, Perú [ROR](#)

* Autor de correspondencia: cagoicocheafa@unitru.edu.pe

Cómo citar: Goicochea Fabián, C. A. (2026). Estrategias didácticas innovadoras y producción oral en inglés como lengua extranjera: revisión de alcance con búsqueda sistematizada (2019–2025). *Lantír Journal of Social Sciences*, 1(1), e003.

<https://doi.org/10.5281/zenodo.22948407>

Copyright © 2026 por el autor. Licenciatario: Lantír Ediciones, Lima, Perú. Artículo de acceso abierto distribuido bajo los términos de la [Licencia Creative Commons Atribución 4.0 Internacional \(CC BY 4.0\)](#).

Resumen

Esta revisión de alcance con búsqueda sistematizada sintetiza la evidencia empírica publicada entre 2019 y 2025 sobre estrategias didácticas innovadoras dirigidas a la producción oral en inglés como lengua extranjera. La versión fue reconstruida a partir de una auditoría bibliográfica del manuscrito inicial y una búsqueda complementaria reproducible en ERIC, portales editoriales y registros DOI, realizada el 24 de septiembre de 2026. Se aplicó un marco PCC: estudiantes de inglés como lengua extranjera, intervenciones didácticas activas o mediadas por tecnología y contextos educativos formales. Se admitieron estudios revisados por pares con resultados observables de desempeño oral o variables afectivas directamente vinculadas con el habla. Nueve estudios verificables integraron la síntesis: gamificación, podcasts, realidad virtual, evaluación entre pares en video esférico, asistentes conversacionales con inteligencia artificial, aprendizaje combinado basado en el enfoque orientado a la producción y entornos digitales de práctica oral. La evidencia muestra resultados favorables, aunque no uniformes. Los efectos más consistentes aparecen cuando la tecnología aumenta el tiempo efectivo de habla, ofrece retroalimentación específica y se articula con tareas comunicativas, rúbricas y secuencias de práctica. Algunos estudios informan mejoras solo en componentes determinados, como gramática y léxico, sin cambios equivalentes en pronunciación, fluidez o disposición para comunicarse. La heterogeneidad de muestras, instrumentos, duración y diseños impide establecer superioridad entre estrategias y desaconseja una estimación agrupada. Se concluye que la innovación eficaz no depende del recurso por sí mismo, sino de su diseño pedagógico, la intensidad de la práctica y la calidad de la evaluación oral.

Palabras clave: producción oral; inglés como lengua extranjera; tecnología educativa; gamificación; pódcast; realidad virtual; inteligencia artificial; revisión de alcance.

Abstract

This scoping review with a systematized search synthesizes empirical evidence published between 2019 and 2025 on innovative teaching strategies aimed at oral production in English as a foreign language. The article was reconstructed through a bibliographic audit of the initial manuscript and a reproducible supplementary search of ERIC, publisher platforms, and DOI records conducted on 24 September 2026. A PCC framework was applied: EFL learners, active or technology-mediated instructional interventions, and formal educational settings. Peer-reviewed studies reporting observable speaking outcomes or affective variables directly related to speaking were eligible. Nine verifiable studies were included, covering gamification, podcasts, virtual reality, peer assessment in spherical video, artificial-intelligence conversational assistants, production-oriented blended learning, and digital speaking environments. Findings were favorable but not uniform. The clearest benefits occurred when technology expanded actual speaking time, provided specific feedback, and was embedded in communicative tasks, rubrics, and sequenced practice. Some studies reported gains in selected components, such as grammar and lexical use, without comparable changes in pronunciation, fluency, or willingness to communicate. Heterogeneity in samples, instruments, intervention length, and study design prevented pooled estimation and precluded ranking strategies. Effective innovation therefore depends less on the tool itself than on pedagogical design, practice intensity, and the quality of oral assessment.

Keywords: oral production; English as a foreign language; educational technology; gamification; podcasting; virtual reality; artificial intelligence; scoping review.

1. Introducción

La producción oral constituye una capacidad central en el aprendizaje del inglés como lengua extranjera porque integra conocimiento lingüístico, procesamiento en tiempo real, manejo del discurso, pronunciación, interacción y regulación afectiva. Hablar no consiste únicamente en reproducir formas gramaticales correctas: implica seleccionar recursos léxicos con rapidez, sostener turnos, responder a un interlocutor, organizar ideas y ajustar el mensaje a una situación comunicativa. Esta complejidad explica por qué estudiantes que obtienen resultados aceptables en comprensión lectora o ejercicios escritos pueden mantener dificultades persistentes al participar en conversaciones, exposiciones o tareas espontáneas.

En muchos contextos escolares y universitarios, las oportunidades de habla son insuficientes. El tamaño del grupo, el tiempo limitado de clase, la prioridad concedida a contenidos evaluables por escrito y el temor a cometer errores reducen el tiempo efectivo de producción de cada estudiante. A ello se añaden la ansiedad lingüística, la percepción de escasa competencia y la falta de interlocutores fuera del aula. El problema no se resuelve simplemente incorporando dispositivos o plataformas: una tecnología puede ampliar las oportunidades de práctica, pero también puede convertirse en un recurso expositivo si no se integra en una tarea que exija producir, negociar y revisar lenguaje.

La literatura reciente emplea el término innovación para designar intervenciones muy distintas. Incluye gamificación, creación de pódcast, simulaciones inmersivas, evaluación entre pares,

asistentes conversacionales, reconocimiento automático del habla y modelos de aprendizaje combinado. Agruparlas bajo una misma etiqueta puede ocultar diferencias importantes. Algunas intervenciones modifican la motivación; otras cambian la frecuencia de la práctica o la calidad del feedback; otras alteran el contexto de interacción. Por ello, el análisis debe identificar el mecanismo pedagógico de cada estrategia y separar los resultados de desempeño de las percepciones o variables afectivas.

La gamificación introduce metas, reglas, retroalimentación inmediata, progresión y elementos de desafío. Su posible contribución a la oralidad no deriva de los puntos o recompensas de manera aislada, sino de la creación de ciclos breves de actuación y feedback. Cuando el alumnado debe resolver problemas, sostener intercambios o repetir una actuación para avanzar, aumenta la práctica distribuida. Sin embargo, una dinámica competitiva mal diseñada puede desplazar la atención desde la calidad lingüística hacia la velocidad o la obtención de recompensas, de modo que la evaluación del desempeño oral debe permanecer alineada con los objetivos lingüísticos.

Los pódcast y las tareas de producción audiovisual combinan planificación, ensayo, grabación, escucha crítica y publicación para una audiencia. Esta secuencia puede favorecer la conciencia fonológica y discursiva porque el estudiante dispone de una representación de su propia actuación y puede revisarla. El valor pedagógico aumenta cuando la grabación culmina en retroalimentación y nueva producción, y disminuye cuando la actividad se limita a leer un guion. En consecuencia, los estudios deben distinguir el producto final de los procesos de interacción oral que conducen a él.

La realidad virtual y los videos esféricos persiguen situar el lenguaje en escenarios más contextualizados. Las simulaciones pueden reducir las restricciones físicas del aula, ofrecer audiencias virtuales y permitir prácticas que serían costosas o difíciles de organizar. No obstante, la sensación de presencia no garantiza aprendizaje. La evidencia debe establecer si el entorno genera una cantidad y calidad de producción oral superior a la condición de comparación y si las mejoras se transfieren a tareas no idénticas a la práctica.

Los asistentes conversacionales basados en inteligencia artificial han ampliado el debate. Su disponibilidad permite mantener intercambios repetidos sin la presión social que algunos estudiantes experimentan frente a docentes o compañeros. Al mismo tiempo, la precisión del reconocimiento, la adecuación del feedback, la protección de datos y la tendencia del estudiante a aceptar correcciones sin examinarlas plantean riesgos. Resulta necesario diferenciar medidas de disfrute, ansiedad y voluntad de comunicarse de las evaluaciones externas de fluidez, precisión, pronunciación o complejidad.

La revisión original reunía trece estudios, pero varias referencias de la tabla no aparecían en la bibliografía o no podían verificarse mediante título, autor, año y fuente. Además, la ecuación de búsqueda no se documentaba por base, no se registraba la fecha de consulta y no se especificaba un instrumento de evaluación crítica. Mantener esas cifras habría producido una apariencia de rigor sin trazabilidad. La presente reconstrucción adopta una revisión de alcance con búsqueda sistematizada: su objetivo es mapear mecanismos, resultados y limitaciones de estudios verificables, no estimar un efecto promedio ni afirmar cobertura exhaustiva de bases cerradas.

La pregunta de revisión fue: ¿qué resultados sobre producción oral y variables afectivas asociadas informan los estudios empíricos de 2019 a 2025 que aplican estrategias didácticas innovadoras en contextos EFL, y qué condiciones de diseño ayudan a interpretar esos resultados? Los objetivos específicos fueron caracterizar participantes, nivel educativo, intervención, comparación e

instrumentos; valorar la calidad del reporte; contrastar resultados positivos, nulos o parciales; y derivar implicaciones prudentes para la enseñanza y para futuras investigaciones.

2. Método

La revisión se diseñó como una revisión de alcance con búsqueda sistematizada y reporte orientado por PRISMA-ScR. Esta denominación responde al tipo de pregunta —amplia, heterogénea y centrada en mapear enfoques— y a las condiciones documentales del expediente. No se presenta como revisión sistemática exhaustiva porque no se conservaron las exportaciones originales de Scopus o Web of Science ni una matriz completa de decisiones. El cambio de clasificación evita atribuir al estudio un nivel de exhaustividad que no puede ser auditado, al tiempo que conserva procedimientos explícitos de elegibilidad, verificación, extracción y evaluación crítica.

El marco PCC delimitó la búsqueda. La población comprendió estudiantes de inglés como lengua extranjera en educación primaria, secundaria o superior. El concepto incluyó intervenciones activas o mediadas por tecnología cuyo diseño generara práctica oral, retroalimentación o interacción. El contexto se restringió a educación formal. Los desenlaces admisibles fueron medidas directas de producción oral —fluidez, precisión, pronunciación, léxico, gramática, inteligibilidad o desempeño global— y variables afectivas estrechamente relacionadas con la conducta de hablar, como ansiedad, confianza o disposición para comunicarse.

Se incluyeron artículos empíricos revisados por pares publicados entre enero de 2019 y diciembre de 2025, en español o inglés, con una descripción identificable de participantes, intervención y resultados. Se aceptaron diseños experimentales, cuasiexperimentales, de método mixto y estudios de intervención con medición pre y post. Se excluyeron revisiones, ensayos teóricos, tesis sin publicación arbitrada, estudios centrados únicamente en comprensión auditiva, lectura, escritura o vocabulario, investigaciones ESL sin una delimitación transferible al contexto EFL y trabajos que no ofrecieran un registro bibliográfico verificable.

La reconstrucción se realizó el 24 de septiembre de 2026. Se auditó primero el corpus del manuscrito mediante coincidencia de autor, año, título, revista y DOI o URL persistente. Después se efectuó una búsqueda complementaria en ERIC, portales editoriales y registros DOI. La ecuación conceptual fue: (EFL OR “English as a foreign language”) AND (speaking OR “oral production” OR “oral proficiency” OR pronunciation OR fluency) AND (gamification OR podcast* OR “virtual reality” OR chatbot OR “artificial intelligence” OR “peer assessment” OR “blended learning”). La sintaxis se adaptó a cada interfaz y se verificó el intervalo 2019–2025 en la ficha editorial.

La búsqueda no se atribuye a Scopus ni Web of Science porque no se dispuso de exportaciones que permitieran documentar fecha, campos, filtros, número de resultados y duplicados. La revisión tampoco usa conteos PRISMA heredados del manuscrito. En su lugar, informa un rastro de auditoría: trece referencias candidatas en la tabla original; siete retiradas por ausencia de referencia completa, falta de correspondencia verificable o desajuste temático; seis referencias iniciales verificadas; y tres estudios adicionales incorporados por búsqueda complementaria. El corpus final quedó compuesto por nueve estudios. Estas cifras describen la reconstrucción editorial y no equivalen a la recuperación exhaustiva de todas las publicaciones existentes.

La selección se realizó en dos etapas. En la primera se examinó título, resumen y registro bibliográfico para comprobar pertinencia y existencia. En la segunda se revisó el texto o la ficha editorial ampliada para confirmar población, intervención y desenlace. Cuando un estudio medía

únicamente ansiedad o disposición para comunicarse se mantuvo solo si la intervención implicaba práctica oral explícita y el resultado permitía comprender una condición afectiva de la participación. Las decisiones se documentaron por referencia para evitar que una cita no verificable reapareciera en la síntesis.

Para cada estudio se extrajeron año, país o contexto, nivel educativo, tamaño de muestra, estrategia, duración cuando estaba disponible, diseño, comparación, instrumento y resultado principal. Se distinguieron resultados de desempeño evaluados por pruebas o rúbricas de los resultados de autoinforme. También se registró si existía grupo de comparación, si se reportaba equivalencia inicial, si la evaluación oral era externa o ciega y si se comunicaban estadísticas suficientes para interpretar el hallazgo.

La evaluación crítica se inspiró en los dominios del Mixed Methods Appraisal Tool: claridad de la pregunta, adecuación del diseño, representatividad o comparabilidad, calidad de la medición, integridad del reporte y correspondencia entre datos y conclusión. No se calculó una puntuación numérica agregada, porque ese procedimiento puede ocultar debilidades diferentes. Se asignó una apreciación global alta, moderada o baja acompañada de una justificación breve. La calidad no se usó para excluir automáticamente, sino para graduar el peso interpretativo de cada estudio.

Debido a la heterogeneidad de intervenciones, poblaciones, instrumentos y estadísticas, se descartó un metaanálisis. La síntesis fue narrativa y se organizó por mecanismo pedagógico: ampliación de la práctica y feedback, producción multimodal, inmersión y contextualización, mediación afectiva y coherencia de la secuencia didáctica. Se evitó convertir la frecuencia de resultados positivos en una estimación de eficacia y se prestó atención a los resultados parciales o nulos.

2.1. Rastro de auditoría del corpus

Tabla 1

Decisiones aplicadas durante la depuración del corpus

Etapa	Decisión documentada	n
Corpus inicial	Referencias candidatas consignadas en la tabla del manuscrito original	13
Auditoría	Retiradas por cita ausente, coincidencia no verificable o desajuste con el desenlace	7
Retenidas	Referencias iniciales comprobadas y pertinentes	6
Complementación	Estudios verificables incorporados mediante búsqueda dirigida	3
Síntesis	Corpus final evaluado narrativamente	9

Nota. Los conteos describen la auditoría editorial y no un flujo de recuperación exhaustivo en bases cerradas.

2.2. Evaluación crítica

La valoración global se interpretó de manera cualitativa. “Alta” indica un diseño y un reporte suficientes para sostener la conclusión principal; “moderada” señala limitaciones que reducen causalidad o generalización, pero no invalidan el hallazgo; “baja” habría indicado deficiencias graves. Ningún estudio fue clasificado como de alta certeza universal, pues todos dependen de su contexto y medición.

3. Resultados

Los nueve estudios incluidos abarcaron educación primaria, secundaria y superior, con una concentración en Asia y presencia de estudios en España, Ecuador e Irán. Predominaron los diseños cuasiexperimentales y de método mixto. Las intervenciones variaron desde secuencias breves con realidad virtual hasta programas de varias semanas con podcast o asistentes conversacionales. Los instrumentos incluyeron pruebas de habla, rúbricas analíticas, calificaciones de jueces, cuestionarios de ansiedad y disposición para comunicarse, entrevistas y registros de participación.

La característica común no fue una tecnología específica, sino la reorganización de la actividad del estudiante. Las intervenciones más prometedoras trasladaron parte del tiempo de clase desde la explicación hacia la producción, permitieron repetir una actuación y proporcionaron información sobre el desempeño. Los resultados fueron menos claros cuando la estrategia se evaluó solo mediante percepción, cuando no existió grupo de comparación o cuando la rúbrica no se describió suficientemente.

En gamificación, Simbaña-Simbaña et al. (2023) trabajaron con cincuenta estudiantes de 13 y 14 años y reportaron mejoras en la comunicación oral después de una intervención con dinámicas gamificadas. El estudio aporta evidencia contextual sobre secundaria, pero su capacidad causal es moderada por la información limitada acerca de asignación, control de variables y tamaño de los efectos. El hallazgo respalda el uso de ciclos de reto y práctica, aunque no permite concluir que cualquier sistema de puntos produzca el mismo resultado.

Los podcast muestran una evidencia más directamente vinculada con la creación oral. Chaves-Yuste (2023) comparó tareas comunicativas convencionales con una intervención de creación y uso de podcast en estudiantes de secundaria. El grupo experimental obtuvo mejores resultados finales en speaking y listening; dentro del grupo experimental también se observaron mejoras significativas en ambas habilidades. El diseño resulta relevante porque la actividad integró contenido social, cooperación, planificación y publicación, de modo que el efecto no puede atribuirse únicamente al formato de audio.

Peng et al. (2025) asignaron setenta estudiantes EFL de nivel intermedio a un grupo con instrucción oral integrada con podcast y a un control con enseñanza habitual. La intervención comprendió diecisiete sesiones durante ocho semanas. El grupo experimental mostró mejoras en habilidades de habla, aprendizaje digital informal y compromiso académico. La combinación de duración, comparación y medición pretest-posttest fortalece la interpretación, aunque la muestra homogénea y el contexto institucional delimitan la generalización.

En realidad virtual, Chien et al. (2020) integraron evaluación entre pares en un entorno de video esférico para el aprendizaje oral. El diseño pone de relieve que la inmersión se vuelve pedagógicamente productiva cuando el estudiante observa criterios, evalúa actuaciones y revisa su propia producción. Ebadi y Ebadijalal (2022) estudiaron Google Expeditions con aprendices iraníes y encontraron efectos favorables en competencia oral y disposición para comunicarse. Ambos trabajos sugieren que la contextualización y el ensayo en escenarios simulados pueden apoyar la oralidad, pero la contribución procede de la tarea y del feedback, no de la presencia visual por sí sola.

Wu y Hung (2022) ofrecen un resultado especialmente útil para evitar conclusiones triunfalistas. En una muestra de 56 estudiantes de sexto grado, el grupo con realidad virtual mejoró

significativamente en gramática y uso léxico respecto del grupo con imágenes, pero no presentó mejoras equivalentes en pronunciación, fluidez, entonación, disposición para comunicarse o autonomía. La intervención, por tanto, benefició componentes específicos y no una competencia oral global. Esta disociación confirma la necesidad de usar instrumentos analíticos y reportar resultados nulos.

Los entornos digitales también pueden actuar sobre variables afectivas. Bashori et al. (2022) analizaron aprendizaje de lenguas mediado por tecnología y ansiedad al hablar. Sus resultados indican que el diseño del entorno, la posibilidad de ensayar y el grado de exposición social modifican la experiencia afectiva. Sin embargo, reducir ansiedad no equivale automáticamente a mejorar fluidez o precisión; ambos tipos de desenlace deben evaluarse por separado.

Zhang et al. (2024) investigaron durante seis semanas un asistente de habla basado en inteligencia artificial con 131 estudiantes universitarios chinos. El estudio encontró cambios favorables en disfrute, ansiedad y disposición para comunicarse. Su aporte principal se sitúa en las condiciones afectivas de la participación y en el control que el aprendiz percibe sobre la práctica. La ausencia de una medida central de desempeño oral comparable obliga a evitar la inferencia de que una mejor disposición produjo necesariamente mayor competencia lingüística.

Cheng (2025) evaluó un modelo combinado basado en el enfoque orientado a la producción con cuarenta estudiantes universitarios. Las producciones iniciales y posteriores fueron calificadas por dos docentes, con alta concordancia entre jueces, y la comparación mostró una mejora significativa. Aunque el diseño careció de grupo control, la explicitación de la rúbrica, la doble calificación y la articulación entre tarea de salida, insumos y evaluación ofrecen una cadena pedagógica plausible. El hallazgo respalda la secuenciación más que una plataforma específica.

En conjunto, la evidencia muestra tres patrones. Primero, las mejoras son más defendibles cuando existe práctica oral reiterada y evaluación directa. Segundo, los efectos pueden ser selectivos: una intervención puede beneficiar léxico o gramática sin modificar pronunciación o fluidez. Tercero, las variables afectivas facilitan la participación, pero no sustituyen la medición del desempeño. No existe base suficiente para ordenar las estrategias de mayor a menor eficacia.

3.1. Características y peso interpretativo de los estudios

Tabla 2

Características metodológicas y peso interpretativo de los estudios

Estudio	Contexto y muestra	Estrategia	Diseño y medida	Resultado principal	Calidad
Chien et al. (2020)	EFL; contexto escolar	SVVR y evaluación entre pares	Cuasiexperimental; desempeño y percepción	Mejoras asociadas con práctica contextualizada y feedback entre pares	Alta
Ebadi y Ebadijalal (2022)	Universitarios iraníes EFL	Google Expeditions VR	Método mixto; prueba oral y WTC	Resultados favorables en competencia oral y disposición para comunicarse	Moderada
Wu y Hung (2022)	56 estudiantes de sexto grado	VR frente a apoyo con imágenes	Dos grupos; prueba oral y cuestionarios	Mejora en gramática y léxico; sin efecto equivalente en pronunciación, fluidez, WTC o autonomía	Alta

Estudio	Contexto y muestra	Estrategia	Diseño y medida	Resultado principal	Calidad
Bashori et al. (2022)	Aprendices EFL	Aprendizaje mediado por tecnología	Intervención; escala de ansiedad	El entorno digital incidió en ansiedad al hablar; no equivale a ganancia de competencia	Moderada
Chaves-Yuste (2023)	78 estudiantes de secundaria, España	Creación y uso de podcast	Cuasiexperimental; pruebas por destreza	Diferencias favorables en speaking y listening	Alta
Simbaña-Simbaña et al. (2023)	50 estudiantes de 13-14 años, Ecuador	Gamificación	Pretest-posttest; comunicación oral	Mejora reportada en componentes de comunicación oral	Moderada
Zhang et al. (2024)	131 universitarios chinos EFL	Asistente de habla con IA	Cuasiexperimental; escalas afectivas	Mayor disfrute y WTC y menor ansiedad; alcance principalmente afectivo	Alta
Peng et al. (2025)	70 mujeres EFL intermedias	Pódcast; 17 sesiones/8 semanas	Dos grupos; pruebas y cuestionarios	Mejora en speaking, aprendizaje digital informal y compromiso	Alta
Cheng (2025)	40 universitarios, China	Aprendizaje combinado basado en POA	Pre-post; dos jueces y rúbrica	Mejora significativa; alta concordancia entre evaluadores	Moderada

Nota. WTC = **willingness to communicate**; SVVR = **spherical video-based virtual reality**; POA = **production-oriented approach**. La valoración se refiere al peso interpretativo del reporte, no a una certificación del estudio.

4. Discusión

La síntesis permite reformular la pregunta habitual sobre qué tecnología funciona. Los estudios no comparan objetos equivalentes y, por ello, resulta más productivo preguntar qué condiciones de diseño transforman un recurso en una oportunidad de aprendizaje oral. La respuesta convergente señala cuatro condiciones: una tarea que obligue a formular significado, tiempo suficiente de producción, retroalimentación interpretable y una nueva oportunidad para actuar. Sin esta secuencia, la innovación corre el riesgo de aumentar novedad o satisfacción sin modificar la competencia.

La cantidad de habla constituye un mecanismo básico, pero no suficiente. Los pódcast, las simulaciones y los asistentes conversacionales pueden ampliar el tiempo de práctica más allá del turno breve disponible en una clase numerosa. No obstante, repetir errores sin orientación puede estabilizarlos. Por ello, la práctica debe combinarse con modelos, criterios y feedback. Las rúbricas analíticas permiten que el estudiante identifique si el objetivo está en fluidez, inteligibilidad, precisión gramatical, amplitud léxica u organización discursiva.

Los resultados de Wu y Hung (2022) ilustran la importancia de evitar indicadores globales. La mejora en gramática y léxico coexistió con resultados no significativos en pronunciación, fluidez e intonación. Si el estudio hubiese comunicado solo una valoración general, esta diferencia habría quedado oculta. Para la docencia, el hallazgo implica que una secuencia de realidad virtual orientada a vocabulario situacional debe complementarse con actividades específicas de ritmo, segmentación, enlace, velocidad y reparación interactiva cuando esos sean los objetivos.

La producción de pódcast reúne varias operaciones valiosas: investigación del contenido, elaboración de guion, ensayo, grabación, escucha de la propia voz y revisión. Su eficacia potencial procede de esta arquitectura. La publicación ante una audiencia real o simulada puede elevar la

responsabilidad comunicativa, pero también inducir una lectura demasiado controlada. Las tareas posteriores deberían incluir preguntas espontáneas, reformulación y conversación para comprobar transferencia desde el producto ensayado hacia el habla menos planificada.

En gamificación, la literatura analizada aconseja separar motivación momentánea de aprendizaje. Un sistema de niveles puede aumentar la participación, pero el objetivo lingüístico debe gobernar la mecánica. La puntuación debería premiar, por ejemplo, el uso de estrategias de reparación, la claridad de la idea, la escucha del interlocutor y la mejora entre intentos, y no solo la rapidez o el número de intervenciones. Además, la competencia pública puede perjudicar a estudiantes con mayor ansiedad; las metas cooperativas y el progreso individual ofrecen alternativas.

La realidad virtual aporta contexto, sensación de presencia y posibilidad de simular escenarios, pero exige decisiones sobre carga cognitiva. Un entorno visualmente complejo puede consumir atención que el aprendiz necesita para planificar el lenguaje. Las actividades iniciales deben familiarizar al estudiante con la interfaz antes de evaluar su producción. También conviene alternar práctica inmersiva con actuaciones frente a interlocutores reales, porque la transferencia es una condición central para valorar el aprendizaje.

La inteligencia artificial conversacional ofrece disponibilidad y baja amenaza social, dos rasgos relevantes para estudiantes que evitan hablar. Sin embargo, la calidad del intercambio depende de reconocimiento de voz, latencia, coherencia de las respuestas y pertinencia de las correcciones. La retroalimentación automática debería tratarse como una fuente que el estudiante contrasta con criterios docentes, no como autoridad infalible. Además, la investigación debe reportar versión del sistema, idioma, configuración y tratamiento de datos, pues estos elementos afectan reproducibilidad y ética.

Los resultados afectivos merecen una interpretación propia. Menor ansiedad y mayor disposición para comunicarse pueden incrementar la probabilidad de que un estudiante participe y persista. Son condiciones habilitadoras, no pruebas directas de dominio lingüístico. Una evaluación sólida debe combinar autoinforme con una tarea oral calificada mediante criterios claros y, cuando sea posible, jueces que desconozcan la condición del participante. Esta triangulación evita confundir entusiasmo por una herramienta con progreso lingüístico.

La heterogeneidad impide una estimación cuantitativa conjunta. Las muestras varían en edad, nivel, lengua materna y experiencia tecnológica; las intervenciones difieren en duración e intensidad; y las pruebas miden componentes distintos. La ausencia frecuente de tamaños de efecto, intervalos de confianza o datos suficientes para calcularlos reduce la comparabilidad. Presentar todos los resultados como mejora significativa produciría una falsa homogeneidad y ocultaría los estudios con efectos parciales.

La concentración geográfica también limita la transferencia. Una estrategia viable en una universidad con dispositivos individuales y conectividad estable puede no serlo en escuelas con acceso intermitente. A la inversa, actividades de pódcast con teléfonos compartidos o gamificación de baja tecnología pueden ser sostenibles si el diseño prevé trabajo cooperativo. Las recomendaciones deben considerar infraestructura, formación docente, tiempo de retroalimentación, accesibilidad y protección de datos.

La evaluación de la producción oral requiere además una correspondencia explícita entre constructo, tarea e interpretación. Una exposición preparada puede revelar organización,

pronunciación y control discursivo, pero ofrece poca información sobre negociación de significado o manejo de turnos. Una entrevista mide interacción, aunque su dificultad depende del interlocutor. Las tareas de descripción, narración y resolución de problemas activan repertorios diferentes. Por ello, los estudios deberían justificar por qué la tarea representa la habilidad que afirman mejorar y utilizar, cuando sea viable, más de una situación comunicativa. La grabación permite una calificación diferida y el control entre jueces, pero no elimina la necesidad de capacitar evaluadores y reportar acuerdo.

También es importante distinguir retroalimentación correctiva, evaluación formativa y puntuación. La corrección señala una discrepancia lingüística; la evaluación formativa relaciona la actuación con criterios y próximos pasos; la puntuación resume el desempeño. Las plataformas suelen ofrecer correcciones inmediatas de pronunciación o gramática, pero pueden no explicar la relevancia comunicativa del error. Un diseño pedagógico sólido combina feedback automático para aspectos frecuentes y observables con retroalimentación docente o entre pares para discurso, adecuación y estrategia interactiva. La revisión posterior debe exigir una nueva actuación: sin ella, la información recibida no se convierte necesariamente en aprendizaje observable.

La equidad es otro criterio de calidad. Las intervenciones digitales pueden beneficiar a quienes disponen de conexión, espacio silencioso, dispositivos recientes y experiencia tecnológica, mientras generan una carga adicional para otros estudiantes. Las diferencias de acento también pueden afectar el reconocimiento automático del habla. Los estudios deberían describir accesibilidad, soporte técnico, pérdidas de participantes y fallos de la herramienta, y analizar si el beneficio se distribuye de forma semejante según nivel inicial. En el aula, deben existir alternativas funcionalmente equivalentes para quien no pueda usar una aplicación y reglas claras sobre conservación de audios, consentimiento y uso de datos.

Para una práctica docente fundamentada, se propone una secuencia de cinco decisiones. Primero, definir el componente oral que se espera mejorar. Segundo, seleccionar una tarea que produzca evidencia observable de ese componente. Tercero, elegir la tecnología solo si reduce una restricción real o aumenta la calidad de la práctica. Cuarto, establecer feedback y oportunidad de revisión. Quinto, evaluar transferencia mediante una tarea distinta de la empleada durante el entrenamiento. Esta lógica evita que la herramienta determine el currículo.

Para la investigación futura, son prioritarios los diseños con grupos comparables, registro previo de hipótesis, descripción de fidelidad de implementación y seguimiento. Los estudios deberían informar asistencia, tiempo efectivo de habla y exposición adicional fuera de la intervención. También deben publicar rúbricas, confiabilidad entre evaluadores, medias y desviaciones o estadísticas equivalentes, tamaños de efecto e intervalos. Estos datos permitirían futuras síntesis cuantitativas y comparaciones más sólidas.

La revisión posee limitaciones propias. La reconstrucción se efectuó sin las exportaciones originales y, por tanto, no permite afirmar exhaustividad ni calcular duplicados de una búsqueda histórica. La consulta se concentró en registros abiertos y portales editoriales; estudios indexados únicamente en servicios cerrados pueden no haber sido recuperados. La evaluación crítica se basó en la información publicada y no en datos proporcionados por los autores. Finalmente, la diversidad de resultados obligó a una síntesis narrativa. Estas limitaciones están declaradas para que el alcance de las conclusiones sea verificable.

5. Conclusiones

La evidencia verificable de 2019 a 2025 indica que las estrategias innovadoras pueden contribuir a la producción oral en EFL, pero sus efectos dependen del diseño pedagógico. Los pódcast, la gamificación, la realidad virtual, la evaluación entre pares, los asistentes conversacionales y el aprendizaje combinado son medios para reorganizar la práctica; ninguno garantiza por sí solo una mejora global. Los resultados más consistentes aparecen cuando la intervención aumenta el tiempo efectivo de habla, incorpora criterios explícitos, ofrece retroalimentación y permite repetir o transferir la actuación.

La síntesis también muestra que los efectos son multidimensionales. Una estrategia puede mejorar gramática y léxico sin producir cambios en pronunciación o fluidez; puede disminuir ansiedad sin demostrar una mejora equivalente en desempeño. Las revistas y los autores deben preservar esos matices, publicar resultados nulos y evitar conclusiones causales cuando el diseño no las sostiene. Para el aula, la recomendación es partir del objetivo lingüístico y utilizar la tecnología únicamente cuando añada una oportunidad concreta de interacción, feedback o contextualización.

La reconstrucción metodológica elevó la trazabilidad del artículo al retirar referencias no verificables, corregir el DOI de Peng et al. (2025), diferenciar desempeño de afecto y sustituir un flujo PRISMA no documentado por un rastro de auditoría explícito. El resultado no pretende cerrar el campo, sino ofrecer un mapa crítico y reproducible que pueda convertirse en una revisión sistemática exhaustiva cuando se disponga de exportaciones completas, cribado independiente y datos comparables.

Referencias

- Bashori, M., van Hout, R., Strik, H., & Cucchiaroni, C. (2022). Effects of technology-mediated language learning on speaking anxiety. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2021.1908244>
- Chaves-Yuste, B. (2023). Podcasts' effects on the EFL classroom: A socially relevant intervention. *Smart Learning Environments*, 10, Article 20. <https://doi.org/10.1186/s40561-023-00241-1>
- Cheng, J. (2025). Blended learning reform in English viewing, listening and speaking course based on the POA in the post-pandemic era. *Frontiers in Education*, 10, 1512667. <https://doi.org/10.3389/feduc.2025.1512667>
- Chien, S.-Y., Hwang, G.-J., & Jong, M. S.-Y. (2020). Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-speaking performance and learning perceptions. *Computers & Education*, 146, 103751. <https://doi.org/10.1016/j.compedu.2019.103751>
- Ebadi, S., & Ebadijalal, M. (2022). The effect of Google Expeditions virtual reality on EFL learners' willingness to communicate and oral proficiency. *Computer Assisted Language Learning*, 35(8), 1975–2000. <https://doi.org/10.1080/09588221.2020.1854311>
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). *Mixed Methods Appraisal Tool (MMAT)*, version 2018. McGill University.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *PLOS Medicine*, 18(3), e1003583. <https://doi.org/10.1371/journal.pmed.1003583>
- Peng, L., Akhter, S., & Hashemifardnia, A. (2025). Podcast-integrated speaking instruction: Enhancing informal digital learning of English, academic engagement, and speaking skills. *Acta Psychologica*, 258, 105158. <https://doi.org/10.1016/j.actpsy.2025.105158>
- Simbaña-Simbaña, E. A., Sarabia-Guevara, D. A., García-Vilema, F. R., & Pico-Poma, J. P. (2023). Gamification for improving oral communication skills in English as a Foreign Language learners. *Cienciamatria*, 9(2), 359–379. <https://doi.org/10.35381/cm.v9i2.1176>

- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Wu, Y.-H. S., & Hung, S.-T. A. (2022). The effects of virtual reality infused instruction on elementary school students' English-speaking performance, willingness to communicate, and learning autonomy. *Journal of Educational Computing Research*, 60(6), 1558–1587. <https://doi.org/10.1177/07356331211068207>
- Zhang, C., Meng, Y., & Ma, X. (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System*, 124, 103259. <https://doi.org/10.1016/j.system.2024.103259>