

<https://doi.org/10.69639/arandu.v13i3.2566>

Evaluación preliminar de Planet Lingua: adecuación del contenido, usabilidad y accesibilidad de un recurso HTML gamificado para la enseñanza de los pronombres personales de sujeto en inglés

Preliminary Evaluation of Planet Lingua: Content Adequacy, Usability, and Accessibility of a Gamified HTML Resource for Teaching English Subject Pronouns

Carmen Luisa Rodríguez Calvache

clrodriguez@ube.edu.ec

<https://orcid.org/0009-0006-9888-4917>

Universidad Bolivariana del Ecuador
Durán, Guayas, Ecuador

Marcos Guerrero Zambrano

mfguerrero@ube.edu.ec

mguerrero@unemi.edu.ec

<https://orcid.org/0000-0002-5617-6836>

Universidad Bolivariana del Ecuador
Universidad Estatal de Milagro
Milagro, Guayas, Ecuador

Ennio Jesus Merida Cordova

ejmeridac@ube.edu.ec

<https://orcid.org/0000-0001-5091-5522>

Universidad Bolivariana del Ecuador
Durán, Guayas, Ecuador.

Artículo recibido: 10 julio 2026- Aceptado para publicación: 16 agosto 2026
Conflictos de intereses: No hay conflictos de intereses

RESUMEN

Objetivo: Valorar preliminarmente la adecuación del contenido, la usabilidad y la accesibilidad de Planet Lingua, un recurso HTML gamificado diseñado para enseñar los pronombres personales de sujeto en inglés en octavo año de Educación General Básica Superior. Métodos: Se realizó un estudio evaluativo de un recurso digital con diseño mixto convergente de predominio cuantitativo (QUAN + qual). Tres registros de evaluación, obtenidos bajo condiciones de inspección no equivalentes, aportaron valoraciones sobre 46 criterios de adecuación, las diez heurísticas de Nielsen y 33 criterios descriptivos basados en WCAG 2.2. Se calcularon estadísticos ordinales, V de Aiken con intervalos de confianza del 95 %, W de Kendall, acuerdo exacto y una puntuación descriptiva ad hoc de accesibilidad. Las observaciones abiertas se analizaron mediante contenido dirigido y se integraron con los resultados cuantitativos en matrices conjuntas. Resultados: De los 44 criterios con información completa, 23 mostraron cumplimiento observado alto, siete requirieron revisión y 14 corrección prioritaria. La concordancia global descriptiva fue $W = 0,62$

y el acuerdo exacto, 40,9 %. Se registraron 35 hallazgos heurísticos de usabilidad, cuatro clasificados como críticos, y 11 criterios de accesibilidad recibieron mayoría de incumplimiento. Los principales riesgos fueron el envío múltiple, la falta de operabilidad completa mediante teclado, el diseño fijo en pantallas estrechas y el control insuficiente del audio. Conclusiones: La versión inicial mostró fortalezas narrativas y de gamificación, pero no aportó evidencia suficiente para considerarla integralmente adecuada ni conforme con WCAG 2.2. La versión mejorada requiere una reevaluación reproducible antes de aplicarse con estudiantes.

Palabras clave: gamificación, tecnología educativa, usabilidad, accesibilidad, inglés como lengua extranjera

ABSTRACT

Objective: To conduct a preliminary assessment of the content adequacy, usability, and accessibility of Planet Lingua, a gamified HTML resource designed to teach English subject pronouns in the eighth year of Ecuadorian upper basic education. **Methods:** An evaluative study of a single digital resource was conducted using a quantitatively driven convergent mixed-methods design (QUAN + qual). Three evaluation records obtained under non-equivalent inspection conditions provided ratings on 46 adequacy criteria, Nielsen's ten usability heuristics, and 33 descriptive accessibility criteria derived from WCAG 2.2. Ordinal descriptive statistics, Aiken's V with 95% confidence intervals, Kendall's W, exact agreement, and an ad hoc descriptive accessibility score were calculated. Open-ended observations underwent directed content analysis and were integrated with the quantitative results through joint displays. **Results:** Of the 44 criteria with complete data, 23 showed high observed compliance, seven required review, and 14 required priority correction. Overall descriptive concordance was $W = .62$, with 40.9% exact agreement. Thirty-five heuristic usability findings were recorded, four of which were classified as critical, and 11 accessibility criteria received a majority of noncompliant ratings. The main risks involved repeated submissions, incomplete keyboard operability, fixed-layout behavior on narrow screens, and insufficient audio controls. **Conclusions:** The initial version showed strengths in narrative design and gamification but did not provide sufficient evidence to be considered adequate overall or conformant with WCAG 2.2. The improved version should undergo a reproducible reevaluation before being tested with students.

Keywords: gamification, educational technology, usability, accessibility, English as a foreign language

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

INTRODUCCIÓN

La enseñanza del inglés como lengua extranjera incorpora recursos digitales interactivos y diseños gamificados para organizar la práctica, ofrecer retroalimentación inmediata y sostener la participación del estudiantado. En contextos de inglés como lengua extranjera o segunda lengua, la gamificación ha utilizado puntos, insignias, niveles, narrativas, retos y tablas de clasificación, con resultados generalmente favorables cuando las mecánicas se alinean con los objetivos educativos y el contexto (Chan & Lo, 2024). Sus beneficios disminuyen cuando los elementos de juego desplazan los objetivos lingüísticos o generan competencia excesiva, distracción o sobrecarga (Zhang & Hasim, 2023). En este marco, la incorporación de Planet Lingua exige una evaluación previa de su contenido, usabilidad y accesibilidad para octavo año de Educación General Básica Superior.

En Ecuador, los juegos digitales han favorecido la práctica de gramática y vocabulario cuando se articulan con propósitos lingüísticos definidos y retroalimentación pertinente (Castillo-Cuesta, 2020). En Educación General Básica Superior de Manabí, una estrategia gamificada activa y cooperativa respondió a dificultades de motivación y actitud identificadas mediante diagnóstico institucional (Pionce-Mendoza et al., 2023). Estos antecedentes muestran que la incorporación de misiones, puntos o insignias no garantiza por sí sola la calidad de un recurso. Antes de utilizar Planet Lingua con estudiantes, es necesario examinar su precisión lingüística, progresión didáctica, claridad de instrucciones, mecanismos de control y condiciones básicas de acceso.

En los antecedentes nacionales consultados, la atención se ha concentrado principalmente en la motivación, la participación y el desempeño, mientras que la evaluación técnica y pedagógica de recursos específicos ha recibido menor atención. En bachillerato ecuatoriano, la selección de recursos y reglas de juego se vinculó con los objetivos curriculares y las características del grupo (Terán Ñacato et al., 2024). En la literatura sobre Duolingo, los efectos de la gamificación móvil variaron según la población, la duración y los indicadores analizados (Shortt et al., 2023). Por ello, la valoración previa de un recurso constituye una fase distinta de la medición posterior de sus efectos sobre el aprendizaje.

La evaluación de Planet Lingua integró tres dimensiones que suelen examinarse por separado: adecuación del contenido, usabilidad y accesibilidad. La evaluación externa de recursos educativos digitales permite identificar debilidades de contenido, diseño pedagógico y funcionalidad antes de su implementación (de la Rosa Gómez et al., 2019). Los marcos fragmentados de validación limitan la interpretación integral de la calidad de los entornos mediados por tecnología (van Haastrecht et al., 2024), mientras que las heurísticas de accesibilidad complementan las comprobaciones técnicas al localizar barreras concretas de percepción, navegación e interacción (Salvador-Ullauri et al., 2020). Un antecedente ecuatoriano

desarrollado parcialmente por el mismo equipo aplicó un esquema integrado a un recurso HTML para enseñar entropía y produjo evidencia diagnóstica y prioridades de rediseño (Valarezo-Chamba & Guerrero-Zambrano, 2026).

En las fuentes consultadas se identificaron estudios sobre los efectos motivacionales y lingüísticos de la gamificación y un antecedente nacional que integró la evaluación del contenido, la usabilidad y la accesibilidad en otro dominio educativo (Valarezo-Chamba & Guerrero-Zambrano, 2026). Dentro del alcance no sistemático de esta revisión narrativa, no se identificaron estudios directamente equivalentes aplicados a un recurso HTML gamificado para la enseñanza de pronombres personales de sujeto en octavo año de Educación General Básica Superior. Los resultados de la gamificación no deben generalizarse automáticamente entre plataformas o poblaciones, porque dependen de la correspondencia entre mecánicas, metas de aprendizaje y contexto (Chan & Lo, 2024). En consecuencia, se formuló la pregunta: ¿qué patrón de adecuación del contenido, usabilidad y accesibilidad presenta Planet Lingua para la enseñanza de pronombres personales de sujeto en inglés en octavo año de Educación General Básica Superior?

El propósito general fue valorar preliminarmente la adecuación del contenido, la usabilidad y la accesibilidad de Planet Lingua para identificar fortalezas, limitaciones y prioridades de mejora antes de su aplicación con estudiantes. El estudio no buscó certificar conformidad formal con estándares ni estimar efectos causales sobre el aprendizaje; se orientó a producir evidencia diagnóstica sobre el contenido, las cuatro misiones, los controles, la retroalimentación y las barreras de acceso.

La pregunta principal se desagregó en cuatro preguntas: 1) ¿qué patrón de adecuación lingüística, didáctica, narrativa, interactiva, gamificada y comunicativa presenta Planet Lingua?; 2) ¿qué problemas de usabilidad se identifican mediante las diez heurísticas de Nielsen y cuál es su severidad?; 3) ¿qué barreras de accesibilidad se identifican mediante criterios agrupados en los principios perceptible, operable, comprensible y robusto?; y 4) ¿qué ajustes deben priorizarse a partir de la integración de los resultados? La contribución del estudio consiste en aplicar e integrar, en un mismo caso educativo ecuatoriano, criterios lingüístico-didácticos, heurísticas de usabilidad y principios descriptivos de accesibilidad, y en traducir sus resultados en prioridades verificables de rediseño.

MATERIALES Y MÉTODOS

Diseño, unidad de análisis y registros de evaluación

Se realizó una investigación aplicada de carácter evaluativo sobre un único recurso digital mediante un diseño mixto convergente de predominio cuantitativo (QUAN + qual) y posicionamiento pragmático. La finalidad fue examinar una versión documentada de Planet Lingua antes de su aplicación con estudiantes, sin manipular variables ni estimar efectos causales sobre el aprendizaje. La unidad de análisis fue Planet Lingua; las unidades de observación fueron

las valoraciones asignadas a cada criterio y las observaciones abiertas asociadas. Las puntuaciones y respuestas se trataron como datos agrupados por registro de evaluación y no como observaciones independientes (Fetters et al., 2013; Kaushik & Walsh, 2019).

Por la naturaleza del estudio no se definió una población ni una muestra de estudiantes. El conjunto analítico estuvo constituido por tres registros identificados como F1, F2 y F3 que corresponde a tres evaluadores expertos en gamificación y en la enseñanza del inglés. El protocolo previó perfiles en didáctica y pedagogía del inglés, tecnología educativa y gamificación aplicada a la enseñanza del inglés, con formación o experiencia comprobable en al menos uno de los dominios centrales, capacidad para ejecutar la actividad y disposición para documentar evidencias. La documentación disponible fue heterogénea: F2 registró experiencia en inglés como lengua extranjera, gamificación y desarrollo web; F1 presentó un dictamen técnico sin ficha individual; y F3 dejó en blanco su caracterización. Esta limitación se mantuvo explícita y no se imputó información no documentada (Yusoff, 2019).

Técnicas e instrumentos

La evaluación integró tres técnicas de inspección externa. La primera utilizó una matriz ad hoc de adecuación del contenido con 46 afirmaciones agrupadas en seis dimensiones: precisión lingüística y cobertura; coherencia didáctica y evaluación; narrativa y audio; mapa, navegación y controles; gamificación y cierre; y comunicación, legibilidad y carga cognitiva. Cada afirmación se valoró en una escala ordinal de cuatro categorías, desde 1 (no cumple) hasta 4 (cumple totalmente); la opción «No observado» se excluyó de los cálculos. La matriz completa está disponible en Figshare (Rodríguez-Calvache & Guerrero-Zambrano, 2026c). La matriz fue creada para este caso y se utilizó con finalidad diagnóstica; no se interpretó como un instrumento previamente validado.

La segunda técnica fue una evaluación heurística de usabilidad basada en las diez heurísticas de Nielsen. El instrumento permitió registrar la heurística, la pantalla o estado, la tarea afectada, el problema, la evidencia, la frecuencia, la severidad y una recomendación. La severidad se clasificó de 0 a 4, desde ausencia de problema hasta problema crítico que impide una tarea esencial o invalida el resultado; no se calculó una puntuación total de usabilidad. La matriz completa está disponible en Figshare (Rodríguez-Calvache & Guerrero-Zambrano, 2026b; Nielsen, 2024).

La tercera técnica utilizó una lista descriptiva de accesibilidad basada en WCAG 2.2, con 33 criterios organizados en los principios perceptible, operable, comprensible y robusto. Las opciones fueron «Sí», «Parcial», «No» y «No aplica». El protocolo contempló pruebas de teclado, foco, zoom del 200 %, ancho de 320 píxeles CSS, espaciado de texto, audio silenciado, reducción de movimiento y dos navegadores cuando fuera posible. Los resultados se interpretaron como descripción de barreras y no como certificación formal de conformidad. La lista completa está disponible en Figshare (Rodríguez-Calvache & Guerrero-Zambrano, 2026a; World Wide Web

Consortium, 2024). Los códigos internos P, O, C y R se utilizaron exclusivamente para organizar la lista y no equivalen a la numeración oficial de los criterios de éxito de WCAG 2.2.

Procedimiento

La fase evaluativa se organizó en cuatro etapas. Primero se consolidó la versión inicial «Planet Lingua Prototipo.html», congelada el 27 de julio de 2026. Entre el 20 y el 27 de julio de 2026, las personas autoras realizaron pruebas funcionales internas en Google Chrome que incluyeron apertura del HTML, registro de participantes, narrativas, mapa, cuatro misiones, respuestas correctas e incorrectas, puntos, insignias, reporte y diploma. Estas comprobaciones corrigieron fallos que impedían aplicar los instrumentos y no se incorporaron como resultados del estudio.

En la segunda etapa se definió una ruta común de familiarización y revisión mediante los tres instrumentos. En la tercera se depuraron los formularios, se verificaron decisiones globales, opciones «No observado» y «No aplica» y evidencias asociadas con puntuaciones bajas, sin imputar datos. La documentación mostró condiciones de evaluación desiguales: F1 correspondió a una evaluación técnica funcional; F2 declaró una revisión estática sin ejecución en navegador; y F3 no consignó dispositivo, navegador ni tecnología de asistencia. En la cuarta etapa se efectuaron los análisis cuantitativo y cualitativo y se integraron los resultados mediante matrices conjuntas que relacionaron puntuación, evidencia, consecuencia, acción de mejora y prioridad. En consecuencia, los análisis de concordancia y recurrencia se interpretaron exclusivamente como exploratorios.

Después de concluir la evaluación se desarrolló una versión mejorada con ajustes derivados de los hallazgos. Este rediseño se consideró un producto posterior y no se incorporó a los cálculos originales de V de Aiken, W de Kendall, evaluación heurística ni puntuación descriptiva de accesibilidad.

Análisis de datos

La extracción, depuración y verificación de los datos, los análisis y la generación de tablas se realizaron mediante Python 3.13.14, pandas 3.0.4, NumPy 2.5.0 y SciPy 1.18.0 (Python Software Foundation, 2026; The pandas development team, 2026; NumPy Developers, 2026; SciPy Developers, 2026). Para la adecuación del contenido se calcularon frecuencias, porcentajes, mediana, rango intercuartílico, mínimo y máximo; la media y la desviación estándar se mantuvieron como descriptores secundarios por el carácter ordinal de la escala.

La favorabilidad de las valoraciones se resumió de forma exploratoria mediante V de Aiken con intervalos de confianza del 95 %. Con tres registros de valoración y cuatro categorías, se aplicaron reglas operativas de interpretación definidas para este análisis: V entre 0,89 y 1,00 indicó cumplimiento observado alto; V = 0,78, revisión; y V igual o inferior a 0,67, corrección prioritaria. Estas reglas no se trataron como umbrales universales. La concordancia se describió mediante W de Kendall con corrección por empates, acuerdo exacto y amplitud media entre

puntuaciones; el valor p se consideró secundario por el número reducido de registros y por sus condiciones heterogéneas (Legendre, 2005; Merino Soto & Livia Segovia, 2009).

Los problemas heurísticos se consolidaron cuando coincidieron en pantalla, tarea y comportamiento observable, y se priorizaron según severidad y recurrencia. Para accesibilidad se asignó 1 a «Sí», 0,5 a «Parcial» y 0 a «No», excluyendo «No aplica»; la puntuación se calculó por registro y principio y se acompañó de un análisis de sensibilidad que asignó 0 a «Parcial». Las observaciones abiertas se examinaron mediante análisis de contenido dirigido; las categorías iniciales correspondieron a las seis dimensiones de contenido, las diez heurísticas y los cuatro principios de accesibilidad. La integración se efectuó mediante matrices conjuntas y distinguió convergencia, complementariedad y divergencia (Assarroudi et al., 2018). El detalle operacional, las ecuaciones y las reglas completas se conservan en el material suplementario metodológico.

Consideraciones éticas

El estudio se limitó a la evaluación preliminar de un recurso educativo digital y no incluyó estudiantes, menores de edad, usuarios con discapacidad ni una intervención pedagógica. Tampoco se recolectaron resultados de aprendizaje, motivación, satisfacción o tiempo de ejecución. Las personas evaluadoras participaron voluntariamente y fueron informadas sobre el propósito del estudio, el uso académico de sus valoraciones y el tratamiento de la información. Para proteger su identidad se emplearon los códigos F1, F2 y F3; no se publicaron datos personales directamente identificables y se informó que la participación no implicaba remuneración, riesgos físicos ni consecuencias académicas o laborales.

RESULTADOS

Descripción de Planet Lingua y disponibilidad del recurso

Planet Lingua es una actividad gamificada interactiva desarrollada para enseñar los pronombres personales de sujeto I, you, he, she, it, we y they en octavo año de Educación General Básica Superior. La versión inicial correspondió al archivo «Planet Lingua Prototipo.html», con presentación visual desarrollada mediante hojas de estilo en cascada y lógica interactiva implementada con JavaScript. El prototipo fue congelado el 27 de julio de 2026; el archivo exacto se conserva sin las mejoras posteriores, no se puso a disposición pública y no se calculó una huella digital SHA-256, limitación que se mantuvo en la trazabilidad del estudio.

La utilización del recurso comprende el registro de cuatro participantes o embajadores y la selección de avatares; una narrativa general y cuatro misiones; síntesis de voz y control de repetición de audio; un mapa con estados de avance; preguntas de selección y retroalimentación para respuestas correctas e incorrectas; puntos, insignias, reporte general y diploma. Las misiones organizan una progresión desde el reconocimiento del referente hasta la aplicación de los pronombres en ejemplos variados.

Después de integrar los hallazgos se desarrolló «Planet Lingua Mejorado.html» mediante HTML, JavaScript y CSS, junto con archivos complementarios para su ejecución. Esta versión se depositó en Figshare como recurso digital citable y no modificó los resultados obtenidos con la versión inicial (Rodríguez-Calvache & Guerrero-Zambrano, 2026d). Enlace al recurso: <https://doi.org/10.6084/m9.figshare.33148544.v1>. El recurso se distribuye bajo una licencia Creative Commons Atribución-NoComercial-SinDerivadas (CC BY-NC-ND).

Como material metodológico complementario se encuentran depositados la matriz de evaluación del contenido, la pertinencia didáctica y la calidad gamificada (Rodríguez-Calvache & Guerrero-Zambrano, 2026c), la matriz de evaluación heurística de usabilidad (Rodríguez-Calvache & Guerrero-Zambrano, 2026b) y la lista de verificación de accesibilidad según WCAG 2.2 (Rodríguez-Calvache & Guerrero-Zambrano, 2026a).

Integridad y alcance del conjunto de datos

La matriz de contenido generó ciento treinta y ocho celdas potenciales, correspondientes a cuarenta y seis criterios por tres registros de evaluación. Se recuperaron ciento treinta y seis puntuaciones válidas. El registro F1 aportó cuarenta y seis puntuaciones, F2 aportó cuarenta y cuatro y F3 aportó cuarenta y seis. Los criterios C4 y C5 fueron registrados como no observados por F2 porque la revisión se limitó al código y no verificó la síntesis de voz en ejecución. En consecuencia, cuarenta y cuatro criterios dispusieron de tres valoraciones comparables y dos quedaron con información incompleta.

La base de usabilidad incluyó treinta y cinco registros originales: dieciocho en F1, doce en F2 y cinco en F3. La base de accesibilidad incluyó noventa y nueve respuestas, con un criterio No aplica en F2 para la compatibilidad entre navegadores. La documentación de los registros fue desigual. F1 se presentó como una evaluación técnica funcional, F2 declaró una revisión estática sin dispositivo ni navegador y F3 no incluyó los metadatos de la persona evaluadora ni de las condiciones de prueba. Esta heterogeneidad limitó la interpretación del acuerdo y obligó a tratar los resultados como evidencia preliminar de tres registros con documentación desigual, no como validación definitiva de un panel homogéneo e independiente.

La Tabla 1 sintetiza la integridad y el alcance de cada registro. F1 aportó información completa en los tres instrumentos, aunque su documento de contenido resumió la evaluación sin una ficha individual de caracterización. F2 documentó su naturaleza estática y registró dos criterios no observados en audio. F3 completó los instrumentos, pero dejó sin consignar los metadatos de la persona evaluadora y de las condiciones tecnológicas.

Tabla 1*Integridad y alcance del conjunto de datos por registro de evaluación*

Registro	Naturaleza	Contenido válido	Hallazgos de usabilidad	Accesibilidad aplicable	Ausencias	Condiciones documentadas
F1	Evaluación técnica funcional	46 de 46	18	33 de 33	Sin N/O ni N/A	Ficha individual no disponible en el resumen
F2	Revisión estática EXP- IA-PL-01	44 de 46	12	32 de 33	C4 y C5 como N/O; ni R6 como por N/A	Sin dispositivo navegador por revisión de código
F3	Matriz complementaria	46 de 46	5	33 de 33	Sin N/O ni N/A	Código, fecha, institución, dispositivo y navegador no consignados

Nota. N/O significa No observado y N/A significa No aplica. La tabla describe la documentación recuperada y no presupone equivalencia entre los tres registros ni independencia entre las valoraciones asociadas.

Distribución de las puntuaciones y resultados por dimensión

Las distribuciones por dimensión mostraron un patrón sistemático de mayor favorabilidad en F3 y menor en F1. Las medianas oscilaron entre 2,00 y 4,00. F1 registró la mediana más baja en precisión lingüística y cobertura, con 2,00, mientras que F3 alcanzó una mediana de 4,00 en las seis dimensiones. Los descriptores ordinales completos por dimensión y registro se presentan en la Tabla S2 del material suplementario.

Valoración exploratoria por ítem mediante el índice V de Aiken

Entre los cuarenta y cuatro criterios completos, veintitrés alcanzaron la categoría operativa de cumplimiento observado alto definida para el estudio, siete quedaron en revisión y catorce requirieron corrección prioritaria. Los criterios B9, C7 y D7 obtuvieron V de 0,33, el valor más bajo observado. B9 examinó la participación de los cuatro embajadores; C7 evaluó la interferencia y el control del audio; y D7 examinó la prevención de envíos múltiples.

La Tabla S3 del material suplementario reúne los veintiún criterios clasificados para revisión o corrección prioritaria. Los IC 95 % fueron amplios, por lo que cada decisión se interpretó junto con las puntuaciones individuales y las observaciones textuales. C4 y C5 no se incluyeron porque solo tuvieron dos valoraciones válidas y requirieron una prueba de audio en ejecución.

. Concordancia entre los registros de evaluación

La concordancia global descriptiva entre los tres registros fue $W = 0,62$ y el acuerdo exacto alcanzó 40,9 %. Las valoraciones conservaron un orden parcialmente semejante, pero no coincidieron en la categoría asignada a la mayoría de los criterios. Como se presenta en la Tabla 2, narrativa y audio registró el mayor valor de W , con 0,83, mientras que comunicación y carga cognitiva obtuvo 0,28.

La precisión lingüística mostró $W = 0,74$, pero solo 14,3 % de acuerdo exacto. Este patrón se produjo porque los registros coincidieron en el orden relativo de varios criterios, aunque utilizaron categorías distintas. La dimensión de comunicación presentó la amplitud media más alta, con 1,20 puntos, y reflejó diferencias sobre legibilidad en pantallas pequeñas, carga cognitiva y representación de estados visuales.

Tabla 2
Concordancia global y por dimensión

Ámbito	Ítems	W	Acuerdo exacto %	Amplitud media	p
Global	44	0,62	40,9	0,93	0,001
Precisión lingüística y cobertura	7	0,74	14,3	1,14	0,038
Coherencia didáctica y evaluación	9	0,58	22,2	1,11	0,086
Narrativa, audio y experiencia audiovisual	6	0,83	66,7	0,33	0,029
Mapa, navegación y controles	8	0,66	62,5	0,75	0,052
Gamificación, puntos, insignias y cierre	9	0,52	55,6	1,00	0,132
Comunicación, legibilidad y carga cognitiva	5	0,28	20,0	1,20	0,504

Nota. W se calculó con corrección por empates. La amplitud media corresponde al promedio de la diferencia entre la puntuación máxima y mínima asignadas a cada ítem. El valor p es secundario y no sustituye la revisión de discrepancias, debido al número reducido de registros, a sus condiciones heterogéneas y a que la independencia de las valoraciones no fue confirmada.

Integración de las observaciones y prioridades de ajuste

La integración cualitativa mostró convergencia en la insuficiente participación de los cuatro embajadores, el control limitado del audio, la necesidad de mejorar la adaptación a pantallas estrechas y la presencia de barreras semánticas. Las divergencias se concentraron en comportamientos verificables del HTML, principalmente la cobertura efectiva de I y you, la retroalimentación contradictoria, el envío múltiple, el momento en que se comunica la regla del 100 % y la proporcionalidad del diploma. La Tabla S4 del material suplementario resume las principales relaciones entre registros y las acciones derivadas.

Evaluación heurística de usabilidad

Los treinta y cinco registros de usabilidad se distribuyeron en tres problemas cosméticos, diecinueve menores, nueve mayores y cuatro críticos. Las heurísticas H3, H5 y H7 concentraron seis registros cada una. H5 y H7 acumularon los cuatro hallazgos críticos: envío múltiple, acceso al mapa mediante teclado y adaptación a pantallas estrechas. La distribución de los registros originales antes de consolidar problemas equivalentes y el número de registros con hallazgos por heurística se presentan en la Tabla S5 del material suplementario.

Después de consolidar descripciones equivalentes, se identificaron seis problemas recurrentes en dos o más registros. La Tabla 3 muestra que el envío múltiple y la falta de operabilidad del mapa por teclado recibieron las severidades máximas más altas. La ausencia de control independiente de la música fue el único problema registrado de forma directa en F1 y F3 y respaldado adicionalmente por las listas de verificación de accesibilidad.

Tabla 3

Problemas recurrentes identificados en dos o más registros

Problema consolidado	H	Registros	Sev. máx.	Consecuencia	Acción
Envío múltiple duplica puntos	H5	F1, F2	4	Invalida puntaje, insignia, reporte y diploma.	Deshabilitar Enviar y opciones tras la primera respuesta.
Mapa no operable por teclado	H7	F1, F2	4	Impide una tarea esencial sin puntero.	Sustituir div por button y gestionar foco y teclado.
Diseño fijo en pantalla estrecha	H7	F1, F2	4	Puede recortar o superponer controles.	Aplicar reflujo, media queries y recorrido vertical.
Control de música insuficiente	H3	F1, F3	3	Reduce control y puede interferir con la voz.	Añadir pausa, silencio, volumen y estado visible.
Criterio de insignia no anticipado	H6	F1, F2	2	La regla se conoce después de completar la misión.	Mostrar el requisito del 100 % antes de iniciar.
Validación de nombres insuficiente	H5	F2, F3	2	Admite entradas poco informativas o duplicadas.	Aplicar trim, longitud mínima y control de duplicados.

Nota. La recurrencia se estableció por equivalencia funcional, no por coincidencia literal. La severidad máxima se conservó para priorizar el rediseño de riesgos esenciales. Dado que la independencia de las valoraciones no fue confirmada, la recurrencia se interpreta de forma exploratoria.

Accesibilidad según los principios WCAG 2.2

Las puntuaciones descriptivas de accesibilidad variaron de forma amplia entre registros. El valor global fue 33,3 % en F1, 46,9 % en F2 y 78,8 % en F3. La puntuación completa por principio y registro se presenta en la Tabla S6 del material suplementario. El principio robusto obtuvo los valores más bajos en F1 y F2, con 14,3 % y 16,7 %, respectivamente, mientras que F3 asignó 64,3 %.

Once criterios recibieron No en al menos dos registros. La Tabla 4 concentra estos criterios, que abarcaron control de audio, marcado del idioma, navegación por teclado, estados programáticos, reducción de movimiento, etiquetas de campos, ayuda, anuncios dinámicos, agrupación de opciones, bloqueo de misiones y adaptación responsiva. P7 fue el único criterio calificado como No en los tres registros.

Tabla 4
Criterios con mayoría de incumplimiento

Código	Criterio	F1	F2	F3	N.º No	Prioridad	Acción
P7	Control independiente de audio	No	No	No	3	Alta	Añadir pausa, silencio y volumen.
P9	Idioma principal y fragmentos en inglés	No	No	Parcial	2	Media	Marcar lang="en" en contenido inglés.
O1	Operabilidad completa mediante teclado	No	No	Parcial	2	Alta	Habilitar Tab, Enter y espacio en todas las funciones.
O4	Estado programático de la opción seleccionada	No	No	Sí	2	Alta	Usar aria-checked o estado equivalente.
O5	Reducción o detención del movimiento	No	No	Parcial	2	Alta	Respetar prefers-reduced-motion.
C5	Etiquetas programáticas de campos	No	No	Parcial	2	Media	Asociar label con cada input.
C8	Ayuda localizable y consistente	No	No	Sí	2	Media	Añadir ayuda persistente y localizable.

Código	Criterio	F1	F2	F3	N.º No	Prioridad	Acción
R2	Anuncio de cambios dinámicos	No	No	Parcial	2	Alta	Incorporar aria-live o role status.
R3	Grupo y estado de opciones	No	No	Parcial	2	Alta	Usar radiogroup y radio o semántica equivalente.
R5	Estado programático de misión bloqueada	No	No	Parcial	2	Alta	Usar disabled o aria-disabled.
R7	Reflujo y adaptación a ventana estrecha	No	No	Parcial	2	Alta	Implementar reflujo y media queries.

Nota. La mayoría de incumplimiento se definió como dos o más respuestas No. Las respuestas Parcial se conservaron para mostrar la dirección de la discrepancia. Los códigos P, O, C y R corresponden a categorías internas del instrumento y no a la numeración oficial de criterios de éxito de WCAG 2.2.

Integración de la adecuación del contenido, la usabilidad y la accesibilidad

La integración identificó problemas que afectaron simultáneamente el contenido, la usabilidad y la accesibilidad. El envío múltiple comprometió la integridad del puntaje y la prevención de errores; el mapa no semántico afectó la navegación, la operabilidad por teclado y la robustez; y el audio sin control independiente afectó la experiencia audiovisual, el control del usuario y la percepción del contenido. La Tabla 5 organiza las acciones resultantes según prioridad, componente, criterio de aceptación y forma de verificación.

La Tabla 5 traduce los resultados en criterios verificables para una nueva versión. Las acciones críticas se consideraron necesarias antes de utilizar Planet Lingua como actividad autónoma o como instrumento con puntaje. Las acciones altas podían implementarse en la misma iteración y verificarse mediante una reevaluación focalizada.

Tabla 5

Integración final y prioridades de rediseño

Prioridad	Acción	Componentes	Criterio de aceptación	Verificación
Crítica	Bloquear el envío múltiple y preservar un único resultado por pregunta.	Botón Enviar, Cada pregunta modifica el puntaje una sola vez y el máximo permanece en 1700.		Prueba funcional con respuestas repetidas.
Crítica	Convertir mapa y trofeo en controles semánticos operables por teclado.	Mapa, misiones, cofre final.	Todos los elementos reciben foco, responden a Enter y espacio y comunican bloqueo.	Recorrido completo solo con teclado y lector de pantalla.

Prioridad	Acción	Componentes	Criterio de aceptación	Verificación
Crítica	Implementar reflujo y adaptación responsiva.	Contenedor, terminal, mapa, reporte y diploma.	No existe pérdida de contenido a zoom 200 % ni a 320 píxeles CSS.	Pruebas en dos navegadores y tres tamaños.
Alta	Corregir cobertura, distractores y retroalimentación.	Banco diecisiete preguntas y mensajes.	de Los siete pronombres son respuestas objetivo; no hay mensajes contradictorios.	Auditoría del banco y prueba de todas las rutas.
Alta	Añadir controles completos de audio y botón de selección de idioma.	Música, voz y botón de repetición.	El usuario puede pausar, silenciar y ajustar; el inglés usa voz o locale apropiado.	Prueba auditiva en dos navegadores.
Alta	Incorporar labels, aria-live, agrupación y estados accesibles.	Registro, opciones, terminal y mapa.	Nombre, función, valor y cambios dinámicos son determinables programáticamente.	Inspección del árbol de accesibilidad y lector de pantalla.
Alta	Hacer explícita la dinámica grupal y las reglas de insignia.	Registro, introducción y misiones.	Cada pregunta asigna rol o turno; el criterio del 100 % aparece antes de iniciar.	Prueba de comprensión con usuarios.
Alta	Reformular el diploma y unificar terminología.	Reporte, diploma y equipos.	El reconocimiento se limita a pronombres personales y usa términos consistentes.	Revisión lingüística y didáctica.
Media	Añadir ayuda, revisión de errores y progreso persistente.	Mapa, reporte y cierre.	La ayuda es localizable y el reporte muestra preguntas falladas y progreso.	Inspección heurística posterior.

Nota. Los criterios de aceptación permiten cerrar cada ajuste mediante evidencia observable. La prueba con usuarios se propone después de corregir las barreras críticas de la versión inicial.

Síntesis evaluativa y análisis de sensibilidad

La síntesis integrada indicó la necesidad de ajustes sustantivos y de una reevaluación focalizada antes de utilizar el recurso de manera autónoma o interpretar sus puntuaciones con

finés evaluativos. Esta decisión se sustentó en catorce ítems con corrección prioritaria, cuatro registros heurísticos críticos y once criterios de accesibilidad con mayoría de incumplimiento. Las fortalezas de la narrativa, el mapa visual, los puntos, las insignias y el reporte no compensaron los riesgos que podían invalidar el puntaje o impedir funciones esenciales.

El análisis de sensibilidad mostró dependencia del registro más favorable y de la ponderación de Parcial. En un escenario de sensibilidad que excluyó F3, dieciséis criterios alcanzaron el máximo posible, siete quedaron en revisión y veintiún criterios permanecieron en corrección prioritaria bajo puntos de corte adaptados a los valores posibles. En accesibilidad, la mediana global disminuyó de 46,9 % a 31,3 % cuando Parcial se puntuó como cero. La Tabla S7 del material suplementario resume estos escenarios.

La variación entre registros impidió interpretar los valores agregados como certificación de adecuación o conformidad. Los resultados respaldaron una evaluación diagnóstica de la versión inicial y orientaron una agenda de rediseño. Como producto posterior, se desarrolló la versión mejorada depositada en Figshare (Rodríguez-Calvache & Guerrero-Zambrano, 2026d). Esta versión deberá someterse a una reevaluación común antes de su aplicación en un estudio con estudiantes.

Producto derivado de la evaluación: versión mejorada

Como producto posterior de la evaluación integrada, se desarrolló una versión mejorada de Planet Lingua. Las modificaciones respondieron a los problemas convergentes de adecuación del contenido, usabilidad y accesibilidad. La actividad y sus archivos complementarios se encuentran disponibles en Figshare (Rodríguez-Calvache & Guerrero-Zambrano, 2026d).

Los cambios implementados se detallan en el Material suplementario 1. Estas acciones describen modificaciones observables en el código y el contenido, pero no constituyen una nueva estimación de adecuación, usabilidad o accesibilidad.

DISCUSIÓN

Interpretación general de los hallazgos

La evaluación mostró un perfil heterogéneo: las fortalezas narrativas, visuales y gamificadas coexistieron con problemas que comprometieron la integridad del puntaje, la interacción y el acceso. Este patrón impide considerar la versión inicial como integralmente adecuada o preparada para un uso autónomo con puntuación. El valor educativo de los elementos de juego depende de su correspondencia con los objetivos lingüísticos, las tareas y la retroalimentación, y no de su presencia aislada (Chan & Lo, 2024).

Los problemas técnicos y de control también produjeron consecuencias pedagógicas. Una mecánica de recompensa deja de apoyar el aprendizaje cuando el puntaje puede alterarse mediante envíos repetidos, cuando las reglas se conocen después de completar la misión o cuando una barrera impide acceder a una tarea esencial. Los fallos técnicos, la explicación insuficiente y los

efectos no deseados de algunas mecánicas pueden reducir los beneficios de la gamificación en contextos EFL/ESL (Zhang & Hasim, 2023).

Adecuación lingüística y educativa

La principal debilidad de contenido se concentró en la representatividad del dominio lingüístico. Los ítems A2 y A3 indicaron que I y you no se trabajaban de forma equivalente como respuestas objetivo y que la distribución favorecía a otros pronombres. La aparición de una forma únicamente como distractor no equivale a una oportunidad de recuperación activa, porque el estudiante puede completar la tarea sin seleccionarla ni justificar su uso. Esta limitación es relevante debido a que el objetivo declarado incluye los siete pronombres personales de sujeto. Los juegos digitales pueden favorecer la gramática y el vocabulario cuando las actividades mantienen objetivos definidos y retroalimentación permanente (Castillo-Cuesta, 2020).

Las puntuaciones de B2, B5, B7 y B9 mostraron que la secuencia propuesta todavía dependía en gran medida del reconocimiento y de la selección cerrada. Los distractores evidentes redujeron la demanda cognitiva, la repetición de referentes favoreció respuestas por familiaridad y la interfaz no organizó la participación de los cuatro embajadores. Las actividades activas y cooperativas pueden responder a necesidades identificadas en Educación General Básica Superior cuando se fundamentan en un diagnóstico institucional (Pionce-Mendoza et al., 2023). En Planet Lingua, esta orientación requiere turnos, roles o una responsabilidad visible por pregunta.

La necesidad de anticipar el criterio del cien por ciento, aclarar la ponderación de las misiones y reformular el diploma también se relacionó con la coherencia entre las reglas del juego y el alcance curricular. Los recursos y las reglas gamificadas deben corresponder con los objetivos y las características del grupo (Terán Ñacato et al., 2024). En Planet Lingua, un diploma que declara dominio general del inglés excede la evidencia generada por diecisiete preguntas sobre pronombres personales de sujeto y puede inducir una interpretación incorrecta del logro.

La distribución de las valoraciones permitió conservar componentes adecuados y localizar criterios que requerían revisión o corrección. La evaluación externa de recursos educativos resulta útil cuando vincula cada puntuación con evidencia localizada y una acción verificable de mejora (de la Rosa Gómez et al., 2019). Debido al conjunto reducido y heterogéneo de registros de evaluación, las estimaciones se interpretaron junto con los intervalos de confianza, las puntuaciones individuales y las observaciones textuales (Yusoff, 2019).

Los valores de W de Kendall y acuerdo exacto describieron una coincidencia parcial en el orden de los criterios, pero no equivalencia en las categorías asignadas. Debido a que las condiciones de inspección no fueron equivalentes y no se confirmó la independencia de todas las valoraciones, estos indicadores se interpretan como descriptores exploratorios y no como evidencia definitiva de acuerdo interevaluador. Las divergencias sobre la retroalimentación contradictoria, el envío múltiple y la comunicación de la regla del cien por ciento describieron comportamientos comprobables del archivo HTML. No pudieron resolverse porque la versión

inicial careció de un identificador técnico inequívoco y las condiciones de inspección fueron heterogéneas; por tanto, deben tratarse como hallazgos pendientes de una nueva inspección reproducible. La interpretación de evidencias de validez requiere relacionar de forma trazable el propósito, la fuente de valoración y el procedimiento de evaluación (van Haastrecht et al., 2024).

Usabilidad heurística y funcionamiento de la interacción

El envío repetido fue el problema funcional más grave, porque podía incrementar puntos y aciertos sin una nueva respuesta válida. Su efecto no se limitó a la facilidad de uso: alteró la correspondencia entre desempeño y resultado, modificó las insignias y comprometió el reporte y el diploma. En consecuencia, la corrección debe preceder a cualquier aplicación en la que el puntaje se utilice para interpretar el aprendizaje o el cumplimiento de las misiones.

La falta de operabilidad del mapa mediante teclado y el diseño fijo en pantallas estrechas también afectan tareas esenciales. Un estudiante que no utiliza un puntero puede quedar detenido en el mapa, mientras que el recorte o la superposición en una ventana reducida puede ocultar controles o contenido. Estos hallazgos deben entenderse como riesgos detectados mediante inspección y no como tasas observadas de fracaso en estudiantes. Después del rediseño se requiere una prueba de tareas con usuarios que registre finalización, errores, necesidad de ayuda y comprensión de las reglas.

Un antecedente nacional desarrollado parcialmente por el mismo equipo evaluó un recurso HTML para la enseñanza de la entropía mediante una combinación de evaluación de contenido, heurísticas y accesibilidad. Ese estudio informó problemas de usabilidad principalmente cosméticos o menores y una puntuación descriptiva de accesibilidad de 81,1 % (Valarezo-Chamba & Guerrero-Zambrano, 2026). Planet Lingua presentó cuatro registros críticos y una mediana global de accesibilidad de 46,9 %. Estos valores no son directamente comparables porque los recursos, los instrumentos y las condiciones de evaluación no fueron idénticos; la comparación se utiliza únicamente para mostrar la transferibilidad potencial del enfoque integrado y no como referencia normativa o evidencia independiente.

Accesibilidad según WCAG 2.2

La amplia variación entre registros y la sensibilidad a la ponderación de Parcial mostraron que el índice descriptivo ad hoc no puede interpretarse como porcentaje de conformidad con WCAG 2.2. El control independiente del audio recibió No en los tres registros y el principio robusto presentó los valores más bajos en F1 y F2. Por tanto, los resultados deben leerse como descripción de barreras y no como certificación normativa.

Las evaluaciones reportaron mecanismos funcionales asociados con las barreras identificadas. Los elementos del mapa implementados como contenedores con activación por clic podían no comunicar adecuadamente nombre, función y estado; los campos sin etiquetas programáticas dependían del marcador de posición; la ausencia de aria-live podía impedir el anuncio de cambios dinámicos; y las opciones sin agrupación semántica podían no exponer su

relación ni su estado de selección. La falta de marcado de idioma podía producir una pronunciación incorrecta en lectores de pantalla, mientras que la ausencia de reducción de movimiento y de reflujo limitaba la adaptación a distintas necesidades. Una evaluación heurística orientada a juegos serios mostró que la revisión humana complementa las comprobaciones automáticas porque localiza barreras de percepción, navegación e interacción que las herramientas técnicas pueden omitir (Salvador-Ullauri et al., 2020).

La amplitud entre los registros no invalida los problemas convergentes, pero sí impide certificar el nivel de accesibilidad. Uno de los registros correspondió a una revisión estática sin navegador y otro no documentó el dispositivo, el navegador ni la tecnología de asistencia. En consecuencia, los resultados describen barreras probables y prioridades de rediseño. La reevaluación debe utilizar una versión congelada, dos navegadores identificados, navegación completa por teclado, zoom de doscientos por ciento, ancho de trescientos veinte píxeles CSS, espaciado ampliado y al menos un lector de pantalla.

Integración, implicaciones y limitaciones

La integración de las tres dimensiones permitió identificar efectos en cascada. El envío múltiple relacionó la integridad del puntaje con la prevención de errores; el mapa no semántico vinculó navegación, teclado y robustez; y el audio sin control afectó la experiencia audiovisual, el control del usuario y la percepción del contenido. Esta meta-inferencia no habría surgido de la lectura aislada de cada instrumento y constituye el aporte principal del estudio: organizar el rediseño por componentes y riesgos interdependientes.

La secuencia de mejora debe comenzar con las funciones que invalidan el resultado o impiden tareas esenciales: bloqueo del envío, operabilidad semántica del mapa y reflujo responsivo. La siguiente iteración debe corregir la cobertura de los siete pronombres, separar las cadenas de retroalimentación, incorporar controles de audio, etiquetas, estados programáticos y anuncios dinámicos. Después corresponde explicitar la dinámica grupal, anticipar las reglas de insignia y limitar el diploma al aprendizaje practicado. Solo una vez verificados estos cambios resulta metodológicamente justificable evaluar satisfacción, usabilidad con estudiantes o efectos sobre el aprendizaje.

La disponibilidad pública de la versión mejorada favorece la trazabilidad del rediseño, pero no demuestra que los problemas hayan desaparecido en condiciones reales de uso. Los ajustes técnicos se describen en el Material suplementario 1 y deberán complementarse con una reevaluación antes de atribuir mejoras en la experiencia de usuario, el acceso funcional o el aprendizaje.

El estudio presenta limitaciones que condicionan el alcance de la discusión. Los tres registros de evaluación se generaron bajo condiciones metodológicamente heterogéneas; uno correspondió a una inspección estática, otro careció de metadatos completos y no se confirmó la independencia de todas las valoraciones. La versión inicial no contó con un identificador técnico

inequívoco ni se depositó en Figshare. La matriz de adecuación fue creada para el caso y no contó con validación previa. El análisis cualitativo no incluyó doble codificación independiente. El índice de accesibilidad utilizó una ponderación ad hoc y las medias se mantuvieron únicamente como descriptores secundarios. No participaron estudiantes, especialistas dedicados a auditoría formal de accesibilidad ni usuarios con discapacidad. Tampoco se midieron aprendizaje, motivación, tiempo de tarea o satisfacción. La participación de quienes desarrollaron el recurso y la ausencia de una verificación independiente de la versión mejorada constituyen fuentes adicionales de posible sesgo. Aunque el análisis de sensibilidad modificó los valores agregados, las prioridades de rediseño se mantuvieron.

La versión mejorada dispone de un identificador persistente en Figshare y deberá someterse al mismo protocolo con evaluadores claramente caracterizados. Una reevaluación focalizada permitirá determinar si las modificaciones reducen los riesgos heurísticos y las barreras de acceso. En una fase posterior, una prueba de usabilidad con estudiantes de octavo año podrá analizar la finalización de tareas, los errores y la comprensión de las reglas antes de estimar cambios en el aprendizaje de los pronombres personales.

CONCLUSIONES

La versión inicial de Planet Lingua presentó fortalezas en la narrativa, la secuencia visual del mapa y la retroalimentación inmediata, pero mostró una adecuación lingüística y educativa parcial: 14 criterios requirieron corrección prioritaria. En usabilidad se identificaron cuatro hallazgos críticos, y en accesibilidad 11 criterios recibieron mayoría de incumplimiento. En conjunto, estos resultados no permiten afirmar una adecuación integral de la versión inicial ni una declaración de conformidad con WCAG 2.2.

Antes de utilizar el recurso de manera autónoma o interpretar su puntaje con fines evaluativos, deben verificarse tres condiciones críticas: un único procesamiento por respuesta, operabilidad semántica completa mediante teclado y reflujo responsivo sin pérdida de contenido. Después deben revisarse la cobertura lingüística, el audio, los estados accesibles, la dinámica grupal y el alcance del diploma.

La evaluación orientó el desarrollo de una versión mejorada con ajustes en la integridad del puntaje, la retroalimentación, la navegación, el diseño responsivo, la cobertura lingüística, la dinámica grupal, el audio, la semántica accesible, la ayuda, el reporte y el diploma. El detalle técnico se presenta en el Material suplementario 1.

La implementación de estas mejoras no modifica los resultados de la versión inicial ni demuestra por sí misma una adecuación integral. La versión mejorada deberá someterse a una reevaluación reproducible antes de estudiar sus efectos sobre el aprendizaje de los pronombres personales de sujeto. Los resultados sugieren la utilidad de integrar la evaluación del contenido,

la usabilidad y la accesibilidad para priorizar riesgos interdependientes antes de evaluar efectos educativos.

Financiación

La investigación no recibió financiación externa.

Conflictos de interés

No hay conflictos de intereses

Contribuciones de autoría

Carmen Luisa Rodríguez-Calvache y Marcos Guerrero-Zambrano participaron conjuntamente en: conceptualización, metodología, software, validación, análisis formal, investigación, curación de datos, diseño de tablas e instrumentos.

Ennio Jesus Merida Cordova participó en la conceptualización, redacción del borrador original, revisión y edición, visualización, supervisión y administración del proyecto. Las tres personas autoras aprobaron la versión final del manuscrito.

Uso de inteligencia artificial

Durante el procesamiento y análisis de los datos se utilizó Gemini de Google como herramienta de apoyo para generar fragmentos de código en Python, posteriormente ejecutados en Google Colab. La herramienta no se empleó para recolectar datos, emitir las valoraciones de los evaluadores, tomar decisiones metodológicas ni formular de manera autónoma la interpretación de los resultados o las conclusiones. El código generado fue revisado, ajustado y ejecutado por las personas autoras, quienes verificaron los resultados obtenidos y asumieron la responsabilidad íntegra por los análisis presentados en el manuscrito.

Disponibilidad de datos, materiales y código

Los instrumentos de evaluación y la versión mejorada de Planet Lingua se encuentran disponibles en Figshare mediante las referencias Rodríguez-Calvache y Guerrero-Zambrano (2026a, 2026b, 2026c, 2026d). La versión mejorada corresponde a un producto posterior y no sustituye la versión inicial que generó los resultados presentados.

REFERENCIAS

- Assarroudi, A., Heshmati Nabavi, F., Armat, M. R., Ebadi, A., & Vaismoradi, M. (2018). Directed qualitative content analysis: The description and elaboration of its underpinning methods and data analysis process. *Journal of Research in Nursing*, 23(1), 42–55. <https://doi.org/10.1177/1744987117741667>
- Castillo-Cuesta, L. (2020). Using digital games for enhancing EFL grammar and vocabulary in higher education. *International Journal of Emerging Technologies in Learning (iJET)*, 15(20), 116–129. <https://doi.org/10.3991/ijet.v15i20.16159>
- Chan, S., & Lo, N. (2024). Enhancing EFL/ESL instruction through gamification: A comprehensive review of empirical evidence. *Frontiers in Education*, 9, Article 1395155. <https://doi.org/10.3389/feduc.2024.1395155>
- de la Rosa Gómez, A., Meza Cano, J. M., & Miranda Díaz, G. A. (2019). Validation of a rubric to evaluate open educational resources for learning. *Behavioral Sciences*, 9(12), Article 126. <https://doi.org/10.3390/bs9120126>
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research*, 48(6, Pt. 2), 2134–2156. <https://doi.org/10.1111/1475-6773.12117>
- Kaushik, V., & Walsh, C. A. (2019). Pragmatism as a research paradigm and its implications for social work research. *Social Sciences*, 8(9), Article 255. <https://doi.org/10.3390/socsci8090255>
- Legendre, P. (2005). Species associations: The Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(2), 226–245. <https://doi.org/10.1198/108571105X46642>
- Merino Soto, C., & Livia Segovia, J. (2009). Intervalos de confianza asimétricos para el índice de validez de contenido: Un programa Visual Basic para la V de Aiken. *Anales de Psicología*, 25(1), 169–171. <https://revistas.um.es/analesps/article/view/71631>
- Nielsen, J. (2024). *10 usability heuristics for user interface design*. Nielsen Norman Group. <https://www.nngroup.com/articles/ten-usability-heuristics/>
- NumPy Developers. (2026). *NumPy* (Version 2.5.0) [Computer software]. <https://numpy.org/>
- The pandas development team. (2026). *pandas* (Version 3.0.4) [Computer software]. <https://pandas.pydata.org/>
- Pionce-Mendoza, K., Véliz-Briones, V., & Mendoza-Bravo, K. (2023). La ludificación en el fortalecimiento del proceso de enseñanza-aprendizaje del idioma inglés en el nivel de básica superior. *Revista Innova Educación*, 5(4), 83–104. <https://doi.org/10.35622/j.rie.2023.05v.005>

- Python Software Foundation. (2026). *Python* (Version 3.13.14) [Computer software]. <https://www.python.org/>
- Rodríguez-Calvache, C., & Guerrero-Zambrano, M. (2026a). *Checklist de accesibilidad para la revisión descriptiva de Planet Lingua según los principios WCAG 2.2* [Online resource]. Figshare. <https://doi.org/10.6084/m9.figshare.33148541.v1>
- Rodríguez-Calvache, C., & Guerrero-Zambrano, M. (2026b). *Matriz de evaluación heurística de la usabilidad de Planet Lingua basada en las diez heurísticas de Nielsen* [Online resource]. Figshare. <https://doi.org/10.6084/m9.figshare.33148538.v1>
- Rodríguez-Calvache, C., & Guerrero-Zambrano, M. (2026c). *Matriz de juicio de expertos para validar el contenido, la pertinencia didáctica y la calidad gamificada de Planet Lingua* [Online resource]. Figshare. <https://doi.org/10.6084/m9.figshare.33148535.v2>
- Rodríguez-Calvache, C., & Guerrero-Zambrano, M. (2026d). *Planet Lingua: recurso gamificado interactivo para el aprendizaje de los pronombres personales sujeto en inglés* [Online resource]. Figshare. <https://doi.org/10.6084/m9.figshare.33148544.v1>
- Salvador-Ullauri, L., Acosta-Vargas, P., González, M., & Luján-Mora, S. (2020). A heuristic method for evaluating accessibility in web-based serious games for users with low vision. *Applied Sciences*, 10(24), Article 8803. <https://doi.org/10.3390/app10248803>
- SciPy Developers. (2026). *SciPy* (Version 1.18.0) [Computer software]. <https://scipy.org/>
- Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2023). Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, 36(3), 517–554. <https://doi.org/10.1080/09588221.2021.1933540>
- Terán Ñacato, M. F., Naranjo Vaca, D. F., Maliza Muñoz, W. F., & Bonilla Tenesaca, J. (2024). Gamificación como estrategia didáctica en el proceso de enseñanza del idioma inglés en el bachillerato general unificado. *Revista Uniandes Episteme*, 11(2), 189–202. <https://doi.org/10.61154/rue.v11i2.3489>
- Valarezo-Chamba, B. S., & Guerrero-Zambrano, M. (2026). Validez de contenido, usabilidad y accesibilidad de un recurso HTML interactivo para la enseñanza de la entropía. *Innova Science Journal*, 4(3), 975–989. <https://doi.org/10.63618/omd/isj/v4/n3/394>
- van Haastrecht, M., Haas, M., Brinkhuis, M., & Spruit, M. (2024). Understanding validity criteria in technology-enhanced learning: A systematic literature review. *Computers & Education*, 220, Article 105128. <https://doi.org/10.1016/j.compedu.2024.105128>
- World Wide Web Consortium. (2024). *Web Content Accessibility Guidelines (WCAG) 2.2*. <https://www.w3.org/TR/WCAG22/>
- Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, 11(2), 49–54. <https://doi.org/10.21315/eimj2019.11.2.6>

Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. *Frontiers in Psychology*, 13, Article 1030790. <https://doi.org/10.3389/fpsyg.2022.1030790>