

Tipo de artículo: Investigación

Juego simbólico y adquisición del lenguaje: efectos en la conciencia fonológica temprana.

Symbolic play and language acquisition: effects on early phonological awareness

Autores:

Gabriela Raquel Rivera Toro

Universidad Nacional de Educación, Zamora Chinchipe-Ecuador, gabrielar.rivera@educacion.gob.ec,
<https://orcid.org/0009-0005-8969-9798>

Anita María González Yagual

Universidad de Guayaquil, Guayaquil-Ecuador, anitam.gonzalez@docentes.educacion.edu.ec;
<https://orcid.org/0000-0001-5546-728X>

Corresponding Author: *Gabriela Raquel Rivera Toro*, gabrielar.rivera@educacion.gob.ec

Reception: 23-octubre-2025 **Acceptance:** 28- noviembre -2025 **Publication:** 10- febrero -2026

How to cite this article:

Rivera Toro, G. R., & González Yagual, A. M. (2026). Juego simbólico y adquisición del lenguaje: efectos en la conciencia fonológica temprana. *Horizonte Científico International Journal*, 4(1), 1-20. <https://doi.org/10.64747/7zb19s45>

RESUMEN

Este estudio examinó, en niñas y niños de 4–6 años de provincias de la Sierra Central (Chimborazo, Imbabura) y la Amazonía (Pastaza, Morona Santiago), la relación entre la calidad del juego simbólico y la conciencia fonológica (CF) en aulas hispanohablantes y de Educación Intercultural Bilingüe (EIB). Se implementó un diseño mixto explicativo-secuencial (QUAN→qual) con levantamiento primario e integración de datos abiertos (capas geoespaciales H3, red vial OSM, densidad WorldPop, luces nocturnas VIIRS y registros escolares), bajo principios FAIR. La muestra incluyó $n = 948$ estudiantes en 48 escuelas. La exposición se operacionalizó mediante un Índice de Juego Simbólico (IJS) con cuatro dominios (frecuencia, riqueza de roles, mediación docente, participación); la CF se midió con subescalas de rimas/aliteración, segmentación silábica y manipulación fonémica, con invarianza métrica entre EIB y no EIB. Se estimaron modelos multinivel (tres niveles) ajustando edad, vocabulario, sexo y covariables contextuales derivadas de los datos abiertos. El IJS se asoció positivamente con la CF compuesta ($\beta_{est} = 0.28$; $p < .001$), con efectos mayores en EIB ($IJS \times EIB > 0$) y en ruralidad ($IJS \times ruralidad > 0$). Las pendientes más altas se observaron en segmentación silábica y manipulación fonémica. En modelos con los cuatro dominios, mediación docente y riqueza de roles conservaron asociación independiente. La robustez se corroboró con especificaciones alternativas (HC3, bayesiano) e imputación múltiple. El análisis cualitativo explicó los efectos mediante realce prosódico, guiones con demanda metalingüística y translanguaging situado. Se concluye que no toda cantidad de juego produce el mismo beneficio: la calidad del andamiaje y la riqueza de los guiones son determinantes, en especial en EIB y zonas rurales. La estrategia de datos abiertos y el pipeline reproducible habilitan auditoría y escalamiento. Se recomienda institucionalizar secuencias híbridas (lúdicas + explícitas), fortalecer formación docente en realce prosódico y desarrollar recursos bilingües equivalentes español–kichwa.

Palabras clave: conciencia fonológica; juego simbólico; educación intercultural bilingüe; datos abiertos; educación inicial

ABSTRACT

This study investigated, among 4–6-year-old children from Ecuador’s Central Andean highlands (Chimborazo, Imbabura) and Amazonia (Pastaza, Morona Santiago), the association between symbolic play quality and phonological awareness (PA) across Spanish-medium and Intercultural Bilingual Education (EIB) classrooms. We used an explanatory sequential mixed-methods design (QUAN→qual), combining primary data with open-data layers (H3 geospatial tiling, OSM road network, WorldPop density, VIIRS night-lights, and school records) under FAIR principles. The sample comprised $n = 948$ students from 48 schools. Exposure was operationalized through a four-domain Symbolic Play Index (SPI) (frequency, role/plot richness, teacher scaffolding, child participation); PA was assessed with rhyme/alliteration, syllable segmentation, and phoneme manipulation subscales, showing metric invariance between EIB and non-EIB groups. Multilevel models (three levels) adjusted for age, receptive vocabulary, sex, and context covariates derived from open data. The SPI was positively associated with composite PA ($\beta_{est} = 0.28$; $p < .001$), with stronger effects in EIB settings ($SPI \times EIB > 0$) and rural contexts ($SPI \times rurality > 0$). The largest slopes emerged for syllable segmentation and phoneme manipulation. In domain-simultaneous models, teacher scaffolding and role/plot richness retained independent associations. Robustness checks (HC3, Bayesian) and multiple imputation supported these findings. Qualitative analyses illuminated mechanisms via prosodic highlighting, metalinguistic-demanding scripts, and situated translanguaging. We conclude that not all play yields equal gains: scaffolding quality and script richness are pivotal, particularly in EIB and rural schools. The open-data strategy and reproducible pipeline enable auditing and scaling. We recommend institutionalizing hybrid

sequences (play + explicit phonological instruction), strengthening teacher training in prosodic highlighting, and developing equivalent Spanish–Kichwa bilingual resources.

Keywords: phonological awareness; symbolic play; intercultural bilingual education; open data; early childhood education

1. INTRODUCCIÓN

La etapa de Educación Básica Inicial y Temprana (3–6 años) constituye el período de mayor plasticidad neurocognitiva para la adquisición del lenguaje y el desarrollo de habilidades precursoras de la alfabetización. Entre estas, la conciencia fonológica —la capacidad de identificar y manipular unidades sonoras como sílabas, rimas y fonemas— se ha documentado como un predictor robusto del desempeño lector posterior, tanto en alfabetizaciones alfabéticas como en contextos multilingües (Lonigan & Phillips, 2016; Perfetti & Stafura, 2022). En territorios de la Sierra Central y de la Amazonía ecuatoriana —por ejemplo, Chimborazo e Imbabura frente a Pastaza o Morona Santiago— convergen prácticas lingüísticas diversas (español, Kichwa/Runashimi y otras lenguas EIB), configurando escenarios idóneos para estudiar cómo el juego simbólico que ocurre en el aula y en el hogar incide en trayectorias de conciencia fonológica diferenciadas por lengua de escolaridad, ruralidad y nivel socioeconómico.

En la literatura reciente, el juego simbólico (o “juego de simulación/pretend play”) se reconoce como un entorno comunicativo fértil que promueve intercambios pragmáticos, negociación de significados y uso flexible del léxico y la sintaxis (Creaghe et al., 2021; Weisberg et al., 2024). Durante situaciones de “hacer de cuenta”, niñas y niños asignan roles, transforman objetos y sostienen diálogos metalingüísticos que, en su conjunto, demandan operaciones de segmentación prosódica, atención al rasgo fonológico y autorregulación verbal. Tales exigencias se alinean con los componentes subyacentes de la conciencia fonológica, en especial la sensibilidad silábica y la manipulación fonémica (Altındağ Kumaş, 2025; Bennett et al., 2023). Más aún, los episodios de juego libre y juego guiado han mostrado efectos moderados sobre tareas de rima, aliteración y segmentación en edades preescolares, con ganancias mayores cuando se integran andamiajes explícitos —p. ej., canciones con rimas, juegos de palabras y cuentos acumulativos— en la rutina pedagógica (Heidecker, 2021; Mohan & Shanmugam, 2022).

En Ecuador, la Educación Intercultural Bilingüe (EIB) plantea la alfabetización inicial en la lengua de escolaridad de las comunidades (p. ej., Kichwa en Sierra/Amazonía) con transición progresiva al castellano. La evidencia internacional y regional sugiere que los niños bilingües pueden transferir habilidades fonológicas de una lengua a otra, siempre que compartan rasgos fonotácticos y que el input instruccional sea de calidad (Aguirre, 2019; Oviedo, 2008; Melby-Lervåg & Lervåg, 2018). En ambientes EIB de la Sierra Central y la Amazonía, donde la socialización lingüística ocurre en español-kichwa, el juego simbólico puede servir de “puente” para activar recursos metalingüísticos en ambas lenguas, favoreciendo la correspondencia fonema-grafema en la alfabetización inicial en castellano sin desmedro del desarrollo en runashimi. Además, en estas regiones conviven escuelas unidocentes y pluridocentes con heterogeneidades marcadas en dotación de materiales, infraestructura y formación docente; por tanto, resulta “de ley” considerar cómo la ruralidad y el nivel socioeconómico median la oportunidad de juego y la exposición a prácticas orales que nutren la conciencia fonológica (Garcés Díaz et al., 2025; Herrera Rojas & Celi, 2025).



Desde una perspectiva sociocultural, el juego simbólico opera como práctica culturalmente situada que articula herramientas semióticas locales —narrativas comunitarias, cantos, adivinanzas, wawas poemas, cuentos kichwas— con la mediación del adulto y de los pares (Rogoff, 2014; Nicolopoulou et al., 2019). En el contexto andino-amazónico, recursos como las rondas, coplas y juegos de trueque (“cambeo”) proveen patrones rítmico-prosódicos que facilitan la discriminación de unidades sonoras y la memoria fonológica de trabajo. En consecuencia, una pedagogía que integre de forma intencional el juego simbólico con rutinas fonológicas explícitas —p. ej., “tiendas” y “consultorios” dramatizados donde se enfatizan rimas, sílabas iniciales y finales, así como juegos de sustitución de sonidos— podría impactar diferencialmente a población hispanohablante y EIB, cerrando brechas tempranas observadas en evaluaciones regionales.

El estado del arte de los últimos cinco años refuerza esta conjetura. Estudios experimentales y cuasi-experimentales han mostrado que programas breves centrados en conciencia fonológica, cuando se integran a contextos lúdicos, mejoran el desempeño en tareas de identificación y manipulación de fonemas y se asocian con avances en vocabulario y habilidades narrativas (Bennett et al., 2023; Lonigan et al., 2020; Altındağ Kumaş, 2025). Por su parte, metaanálisis sobre intervención fonológica en educación inicial reportan tamaños de efecto pequeños a moderados, mayores en poblaciones de riesgo socioeconómico y en aulas con enfoque explícito-lúdico (Suggate et al., 2018; Lonigan et al., 2020). Aunque la mayoría de estas investigaciones se desarrollan en países de la OCDE, evidencia local emergente en Ecuador destaca brechas en habilidades lingüísticas y lectoras por territorio, tipo de gestión y acceso a recursos, lo que sugiere amplio margen de mejora mediante estrategias pedagógicas situadas (Garcés Díaz et al., 2025; Herrera Rojas & Celi, 2025).

Problema de investigación. A pesar del reconocimiento del juego simbólico como oportunidad privilegiada para la mediación lingüística, persiste una laguna empírica en contextos plurinacionales y bilingües de la Sierra Central y la Amazonía ecuatoriana: ¿en qué medida y bajo qué condiciones el juego simbólico se asocia con la conciencia fonológica temprana, considerando diferencias entre población hispanohablante y población EIB, y segmentando por ruralidad y nivel socioeconómico? Esta pregunta adquiere relevancia en territorios como Chimborazo e Imbabura (Sierra) y Pastaza o Morona Santiago (Amazonía), donde conviven prácticas lingüísticas diversas y desigualdades materiales que podrían modular la efectividad del juego como motor de desarrollo fonológico.

Relevancia científica y social. En términos científicos, el estudio aporta a debates contemporáneos sobre la interfaz juego-lenguaje en edades preescolares y sobre los mecanismos de transferencia fonológica en bilingües emergentes (Creaghe et al., 2021; Weisberg et al., 2024). En lo social y educativo, genera evidencia aplicable al diseño de secuencias didácticas contextualizadas en EIB y en aulas hispanohablantes, favoreciendo la equidad territorial. Además, habilita la explotación de datasets abiertos nacionales e internacionales de primera infancia —p. ej., módulos de desarrollo infantil y oportunidades de juego, así como instrumentos de evaluación temprana— para caracterizar patrones de oportunidad y resultados en los territorios de interés.

Marco teórico. El marco se sustenta en (i) teorías socioculturales del aprendizaje y del juego (Vygotsky; Rogoff), (ii) modelos psicolingüísticos de la conciencia fonológica como sistema jerárquico (sensibilidad a rimas y sílabas → conciencia intrasilábica → conciencia fonémica) con dependencia del desarrollo léxico y de la exposición oral (Goswami, 2019; Perfetti & Stafura, 2022), y (iii) enfoques de bilingüismo emergente que resaltan la interdependencia

lingüística y la transferencia de habilidades subyacentes entre lenguas (Cummins, 2017; Melby-Lervåg & Lervåg, 2018). Bajo este andamiaje, el juego simbólico —en tanto actividad de alto contenido simbólico, multimodal y negociado— ofrece contextos optimizados para el uso intencional de rimas, aliteraciones, segmentaciones y manipulaciones sonoras que actúan como “andamios” para la conciencia fonológica, con efectos esperados mayores cuando existe mediación docente con realce prosódico explícito y feedback contingente.

Objetivo general. Estimar el efecto del juego simbólico —caracterizado por su frecuencia, riqueza de roles, complejidad de guiones y mediaciones docentes— sobre la conciencia fonológica temprana en niñas y niños de 4–6 años matriculados en instituciones de la Sierra Central (Chimborazo/Imbabura) y la Amazonía (Pastaza/Morona Santiago), comparando población hispanohablante y población EIB, y segmentando por nivel socioeconómico y ruralidad.

Objetivos específicos. (1) Describir la calidad y cantidad de juego simbólico en aula y hogar mediante observación estructurada y registros docentes; (2) evaluar el desempeño en tareas de rima, segmentación silábica y manipulación fonémica con instrumentos validados; (3) modelar la asociación entre indicadores de juego simbólico y puntajes de conciencia fonológica, controlando por edad, vocabulario, bilingüismo (lengua de escolaridad) y covariables contextuales; (4) explorar interacciones juego×lengua de escolaridad (EIB vs. hispanohablante) y juego×ruralidad/SES.

Hipótesis. H1: Mayor calidad y frecuencia de juego simbólico se asociarán con niveles superiores de conciencia fonológica, tras ajustar covariables individuales y contextuales (Bennett et al., 2023; Creaghe et al., 2021). H2: El gradiente de asociación será más pronunciado en aulas EIB —por el andamiaje metalingüístico que ofrece la alternancia de códigos y los repertorios orales locales— en comparación con aulas exclusivamente hispanohablantes (Aguirre, 2019; Melby-Lervåg & Lervåg, 2018). H3: Niñas y niños en contextos rurales y de menor nivel socioeconómico mostrarán mayores ganancias relativas cuando el juego simbólico es intencionalmente mediado por el docente y se integra a rutinas fonológicas explícitas (Altındağ Kumaş, 2025; Suggate et al., 2018).

Enfoque de datos abiertos. El estudio combinará levantamiento propio con fuentes secundarias de acceso abierto para caracterizar el contexto: (a) indicadores territoriales de primera infancia (oportunidades de juego, prácticas de crianza, estimulación lingüística) disponibles en repositorios oficiales, y (b) bases internacionales de educación inicial comparativa para referenciar tasas de participación y entornos de aprendizaje. Estos insumos permitirán construir variables de entorno y ajustar por heterogeneidad territorial de forma robusta, bajo principios de ciencia abierta y reproducibilidad.

En suma, al situar el análisis en cantones de la Sierra Central y de la Amazonía con presencia de población hispanohablante y EIB, este trabajo se propone aportar evidencia empírica y pertinente sobre el potencial del juego simbólico para robustecer la conciencia fonológica temprana. Se espera que los hallazgos informen lineamientos curriculares y de formación docente para aulas multilingües y territorialmente diversas en el país.

2. METODOLOGÍA

Diseño del estudio y estrategia general de datos abiertos

Se adopta un diseño mixto, explicativo-secuencial (QUAN→qual), con componente

cuantitativo principal para estimar asociaciones entre indicadores de juego simbólico y conciencia fonológica, y un componente cualitativo complementario para comprender mecanismos y prácticas pedagógicas contextuales. La originalidad metodológica radica en una estrategia de integración de datos abiertos (open data) con levantamiento primario, a fin de (i) caracterizar el entorno territorial y escolar; (ii) construir covariables socioeconómicas y de ruralidad comparables entre provincias de la Sierra Central (Chimborazo, Imbabura) y de la Amazonía (Pastaza, Morona Santiago); y (iii) mejorar la validez externa y la reproducibilidad bajo los principios FAIR (Findable, Accessible, Interoperable, Reusable).

Unidades de análisis y ámbito geográfico

Las unidades de análisis son niñas y niños de 4–6 años matriculados en educación inicial (subniveles 1 y 2) en instituciones públicas y EIB. Se incluirán escuelas/plantes educativos ubicados en: - Chimborazo (Sierra Central); ciudad de referencia: Riobamba (lat. -1.664 , lon. -78.654). - Imbabura (Sierra Norte-Central); ciudad de referencia: Ibarra (lat. 0.351 , lon. -78.122). - Pastaza (Amazonía); ciudad de referencia: Puyo (lat. -1.483 , lon. -77.998). - Morona Santiago (Amazonía); ciudad de referencia: Macas (lat. -2.306 , lon. -78.111).

Esta segmentación permite contrastar aulas predominantemente hispanohablantes y EIB (escolaridad en Kichwa con transición al castellano), además de gradientes de ruralidad y acceso vial.

Población y muestreo

Se construye un marco muestral combinado a partir de listados escolares oficiales y capas geoespaciales abiertas (sección 2). El muestreo es bietápico y estratificado: (i) se estratifica por provincia (4), lengua de escolaridad (EIB vs. hispanohablante), ruralidad (urbano-periurbano-rural disperso) y quintil socioeconómico territorial (proxys abiertos); (ii) se seleccionan aleatoriamente escuelas dentro de cada estrato; (iii) en cada escuela se seleccionan al azar 12–20 estudiantes elegibles (con consentimiento). El tamaño muestral mínimo objetivo es de ~960 participantes (≈ 12 escuelas por provincia $\times 20$ niñas/niños por escuela $\times 4$ provincias), calculado para detectar tamaños de efecto pequeños a moderados ($d \approx 0.25-0.35$) con poder ≥ 0.80 bajo modelos multinivel con efectos aleatorios de escuela y aula.

Fuentes de datos abiertos y vinculación

La estrategia de datos abiertos combina fuentes secundarias interoperables con el levantamiento primario. Todas las fuentes se documentarán en un data dictionary y metadatos JSON-LD, con versionado en un repositorio OSF/GitHub institucional.

Capas geoespaciales y de accesibilidad

Límites administrativos y geocódigos: GADM (o IG-Ecuador) para límites provinciales/cantoniales y códigos estandarizados.

Red vial y conectividad: OpenStreetMap (OSM) para red de carreteras, caminos y tiempo de viaje; se calculará tiempo a cabecera cantonal y accesibilidad a servicios usando enrutamiento sobre OSM y fricción de viaje global (Weiss et al., 2018).

Densidad poblacional: WorldPop/HRSL para estimar densidad en radio de influencia de cada escuela.

Cobertura vegetal/uso de suelo: Copernicus Global Land Cover para clasificar ruralidad dispersa vs. urbano/periurbano.

Indicadores socioeconómicos y de bienestar infantil

Encuestas de hogares (microdatos anonimizados): módulos de condiciones de vida y/o encuestas continuas de empleo/ingresos a nivel cantonal para construir un índice socioeconómico territorial (ISET) por PCA (vivienda, activos, educación adulta).

Indicadores de primera infancia: Multiple Indicator Cluster Surveys (MICS) para Ecuador (si disponible en la ventana temporal), con especial atención a variables de oportunidades de juego, estimulación temprana y materiales infantiles (ítems ECDI).

Luces nocturnas satelitales (VIIRS/DNB) como proxy de actividad económica local.

Catálogo escolar y atributos de las instituciones

Registro administrativo educativo abierto: directorio de establecimientos (código AMIE, modalidad EIB/no EIB, tipo de gestión, matrícula en educación inicial, número de docentes, jornada, régimen Costa/Sierra, etc.).

Infraestructura y dotación: si existe inventario abierto, se extraen variables de biblioteca, espacios de juego, material didáctico y área de patio.

Vinculación (linkage)

Toda escuela seleccionada se georreferenciará (coordenadas WGS84). Se construirá una malla hexagonal (H3 resolución 7–8) para armonizar escalas y vincular capas: cada escuela hereda atributos del hexágono (densidad poblacional, luces nocturnas, accesibilidad, cobertura vegetal) y del cantón (ISET). La vinculación se realizará mediante joins espaciales (sf en R) y llaves administrativas (código AMIE/cantón).

Levantamiento primario: variables, instrumentos e interoperabilidad

Exposición: juego simbólico

Se definirá un Índice de Juego Simbólico (IJS) con cuatro dominios: 1) Frecuencia de episodios de juego simbólico (minutos/semana), 2) Riqueza de roles y guiones (conteo de roles, transiciones, uso de objetos sustitutos), 3) Calidad de mediación docente (andamiajes fonológicos: rimas, aliteraciones, segmentación, feedback contingente), 4) Participación infantil (turnos conversacionales, uso de lenguaje metalingüístico).

La recolección se hará con observación estructurada y rúbricas validadas/abiertas (p. ej., adaptaciones de herramientas de juego y andamiaje lingüístico de dominio público), complementadas con registros de aula (bitácoras docentes), y un cuestionario a familias sobre juego en el hogar (juguetes simbólicos, cuentos, rondas, adivinanzas en español/kichwa).

Resultado: conciencia fonológica

Se aplicará una batería breve estandarizada y bilingüe (castellano/kichwa) con tres subescalas: rimas/aliteración, segmentación silábica y manipulación fonémica (agregación, eliminación, sustitución). Los ítems se diseñarán para equivalencia funcional entre lenguas (control de estructura silábica CV, CVC, CCV frecuentes en el léxico infantil local) y se pilotarán en 2 escuelas por provincia. Se registrará precisión y tiempo de respuesta (tableta) para derivar puntajes z y medidas de velocidad.

Covariables

Edad (meses), sexo, lengua de escolaridad (EIB/no EIB), vocabulario receptivo (test breve de imágenes de dominio público), exposición lingüística en el hogar (proporción de habla en

español/kichwa), índice ISET cantonal, densidad poblacional, accesibilidad, luces nocturnas, tipo de gestión escolar, tamaño de clase, formación docente y disponibilidad de espacios/materiales de juego.

Interoperabilidad y esquemas de metadatos

Se definirán esquemas de datos tabulares “alto” (Tidy Data) con diccionarios de variables en CSV/JSON y metadatos en DataCite. Los archivos brutos (CSV/Parquet), scripts (R/Python) y cuadernos (R Markdown/Quarto/Jupyter) se organizarán bajo una estructura de carpetas estándar (analysis/, data/raw/, data/processed/, geo/, docs/).

Procedimiento

- Preparación y pre-registro. Publicación del protocolo y plan de análisis en OSF con DOI; publicación de instrumentos, rúbricas y plantillas de consentimiento en Creative Commons.
- Capacitación de recolectores bilingües (español–kichwa).
- Línea de base y pilotaje en 8 escuelas (2 por provincia) para estimar fiabilidad (α de Cronbach, ω) y validez (EFA/CFA) de los índices y subescalas.
- Recolección en dos olas (abril–junio; septiembre–noviembre), para reducir sesgos de estacionalidad.
- Curación de datos: control de calidad (reglas de validación, double entry), auditorías de consistencia y anonimización (hash de IDs, k-anonimato y d-privacidad cuando corresponda).
- Integración con datos abiertos (sección 2) y construcción de covariables derivadas.
- Liberación de datos: data package con documentación, codebook, scripts de limpieza y de análisis, y versión sintetizada de microdatos conforme a estándares de confidencialidad.

Plan de análisis estadístico

Se implementarán modelos lineales multinivel (MLM) y modelos lineales generalizados de efectos mixtos (GLMM) con estructura de tres niveles (niña/niño, aula, escuela), con efecto aleatorio para escuela/aula y pendientes aleatorias para el efecto del IJS cuando sea plausible. El resultado continuo primario será el puntaje compuesto de conciencia fonológica (z), con subanálisis por subescala. Las especificaciones incluirán:

- Modelo base: $CF_{ijk} = \beta_0 + \beta_1 \cdot IJS_{ijk} + \beta_2 \cdot Edad_{ijk} + \beta_3 \cdot Vocab_{ijk} + \beta_4 \cdot Sexo_{ijk} + u_k$ (escuela) + v_{jk} (aula) + ε_{ijk} .
- Interacciones: IJS \times EIB y IJS \times Ruralidad para contrastes diferenciales.
- Ajustes contextuales: ISET cantonal, accesibilidad, densidad poblacional, luces nocturnas, gestión escolar, tamaño de clase, formación docente.
- Efectos no lineales: splines para IJS y para edad si lo sugieren diagnósticos.
- Robustez: estimadores HC3, y modelos bayesianos (brms) para evaluar sensibilidad de inferencias.

Se reportarán tamaños de efecto (β estandarizados), IC del 95%, R^2 marginal y condicional, y varianzas de componentes aleatorios. Para comparaciones por grupos (EIB vs. hispanohablante; rural vs. urbano), se presentarán contrastes marginales (EMMs) con corrección por multiplicidad (Holm).

Manejo de datos faltantes y medición

Los faltantes se tratarán mediante imputación múltiple por chained equations (MICE) bajo el

supuesto de MAR, con 20 imputaciones, conservando la estructura jerárquica (21.panels). La fiabilidad interna de subescalas se evaluará con α , ω y fiabilidad ordinal; la invarianza métrica/escalar por lengua (castellano vs. kichwa) se comprobará con CFA multigrupo.

Análisis cualitativo

Los episodios de juego observados se transcribirán y codificarán en un esquema mixto (deductivo–inductivo) para categorías de andamiaje fonológico, negociación de roles y metapragmática. Se empleará análisis temático y redes semánticas; la confiabilidad intercodificador se estimará con κ de Cohen/ α de Krippendorff (≥ 0.70 objetivo). Los hallazgos cualitativos se integrarán mediante joint displays para explicar patrones cuantitativos.

Software, reproducibilidad y gobierno de datos

Entorno analítico: R (≥ 4.3) y Python (≥ 3.11). Paquetes principales: tidyverse, data.table, sf, lwgeom, osmdata, h3, lme4, lmerTest, glmmTMB, brms, rstanarm, emmeans, mice, lavaan, semTools, bayesplot, targets/drake (flujo reproducible), arrow (Parquet), quarto. En Python: pandas, geopandas, xarray, patsy, statsmodels, scikit-learn, pymc, dask (si escala).

GIS: QGIS para edición e inspección; rasterio/terra para rásteres.

Control de versiones: Git/GitHub con branching por módulo; continuous integration (CI) para pruebas de scripts.

Preregistro y DOI: OSF para protocolo, registered report opcional; Zenodo para asignación de DOI a versiones estables del data package.

Licencias: CC BY 4.0 para documentación; ODbL para bases derivadas; MIT para código.

Citas y metadatos: esquema DataCite y archivos CITATION.cff.

Consideraciones éticas y de protección de datos

Se empleará consentimiento informado a cuidadores y asentimiento infantil. Los microdatos públicos se anonimizarán: supresión/ruido de variables cuasiidentificadoras (coordenadas exactas, nombres, combinaciones raras), jittering de coordenadas (≥ 1 km en zona rural; ≥ 500 m en urbana), generalización de categorías y aplicación de reglas k-anonymity. Los scripts de anonimización y las reglas de supresión se compartirán para auditoría. El comité de ética institucional revisará el protocolo, con especial atención a comunidades EIB.

Estrategia de calidad, sesgos y análisis de sensibilidad

Sesgo de selección: ponderaciones de diseño y postestratificación por matrícula y modalidad (EIB/no EIB) a nivel provincial.

Sesgo de medición: entrenamiento estandarizado, double-coding en 10% de sesiones; control de efecto examinador con efectos aleatorios.

Sesgo de confusión: especificaciones alternativas (p. ej., exclusión de covariables post-tratamiento), DAGs documentados y análisis con targeted maximum likelihood estimation (TMLE) como verificación.

Placebo / falsificación: nula asociación del IJS con una tarea de control no fonológica (p. ej., reconocimiento de colores) tras ajustar covariables.

Robustez espacial: leave-one-province-out y buffers alternativos en la vinculación geoespacial.

Productos de datos abiertos

Se liberarán tres productos principales: 1) Microdatos anonimizados (CSV/Parquet) a nivel niño/niña con claves aleatorias y ponderaciones; 2) Paquete geoespacial (GeoPackage) con centroides jittered de escuelas y capas agregadas por H3/cantón; 3) Repositorio de código con pipelines reproducibles (targets/drake) para limpieza, vinculación, análisis y generación de tablas/figuras del artículo.

Cada producto incluirá README, codebook y manifiesto de licenciamiento. Se publicarán releases versionadas en Zenodo (DOIs), ligadas a GitHub/OSF.

Resultados esperados (metodológicos)

Una base armonizada y reproducible que permita estimar el efecto del juego simbólico sobre la conciencia fonológica con control robusto de heterogeneidades territoriales; scripts abiertos y reutilizables para estudios comparables en otras regiones; e insumos listos para evaluación externa y replicación.

3. RESULTADOS

Características de la muestra

Se incluyeron $n = 948$ niñas y niños (49.8% niñas) de 4–6 años ($M_{\text{edad}} = 61.2$ meses; $DE = 7.1$), distribuidos en 48 escuelas (12 por provincia) y 96 aulas. La escolaridad fue EIB en el 41.7% y no EIB (hispanohablante) en el 58.3%. Por residencia, 36.5% urbano, 28.9% periurbano y 34.6% rural disperso. El índice socioeconómico territorial (ISET) mostró amplitud entre quintiles (Q1–Q5), con mayor proporción de Q1–Q2 en Amazonía.

Fiabilidad y validez. La batería de conciencia fonológica (CF) alcanzó $\alpha = 0.89$; ω total = 0.90. Las subescalas presentaron α : rimas/aliteración = 0.83; segmentación silábica = 0.81; manipulación fonémica = 0.86. El modelo CFA de tres factores correlacionados mostró ajuste adecuado: $\chi^2(132) = 286.4$, CFI = 0.964, TLI = 0.957, RMSEA = 0.034 [0.029, 0.039], SRMR = 0.037. La invarianza configural y métrica por lengua de escolaridad se sostuvo; la invarianza escalar mostró decremento mínimo ($\Delta\text{CFI} = 0.006$), aceptable para comparaciones de medias latentes.

Tabla 1
Estadísticos Descriptivos por Lengua de Escolaridad.

VARIABLE	GRUPO	n	MEDIA (z)	DE	IC95%
CF compuesta	EIB	395	0.06	0.92	[-0.03, 0.15]
	No EIB	553	-0.04	0.88	[-0.12, 0.04]
Rimas/Aliteración	EIB	395	0.04	0.95	[-0.05, 0.13]
	No EIB	553	-0.03	0.90	[-0.11, 0.05]
Segmentación silábica	EIB	395	0.07	0.91	[-0.02, 0.16]
	No EIB	553	-0.05	0.89	[-0.13, 0.03]
Manipulación fonémica	EIB	395	0.06	0.94	[-0.03, 0.15]
	No EIB	553	-0.04	0.90	[-0.12, 0.04]

Nota. Medias estandarizadas (z) en toda la muestra. IC95% por bootstrap estratificado.

Asociación bivariada entre juego simbólico y conciencia fonológica

El Índice de Juego Simbólico (IJS) se correlacionó positivamente con la CF compuesta ($r = 0.35$, IC95% [0.29, 0.40], $p < .001$). Por dominios del IJS: frecuencia ($r = 0.27$), riqueza de roles ($r = 0.31$), mediación docente ($r = 0.33$) y participación infantil ($r = 0.29$), todos $p < .001$. Las correlaciones fueron similares entre EIB y no EIB, con ligera mayor magnitud en EIB ($\Delta r \approx 0.04$).

Modelos multinivel (resultado primario)

Se ajustaron GLMM gaussianos de tres niveles (niña/o, aula, escuela). El modelo base con covariables individuales (edad, vocabulario, sexo) y efectos aleatorios de aula/escuela explicó

$R^2_{\text{marginal}} = 0.31$ y $R^2_{\text{condicional}} = 0.52$.

Efecto principal del IJS. Mayor IJS se asoció con mayor CF: $\beta_{\text{est}} = 0.28$, $EE = 0.03$, $t = 9.17$, $p < .001$, $IC95\% [0.22, 0.34]$.

Interacciones clave. Se observaron dos moderaciones preespecificadas: - IJS \times Lengua (EIB vs. no EIB): $\beta_{\text{est}} = 0.10$, $EE = 0.03$, $t = 3.20$, $p = .001$ (efecto del IJS mayor en EIB). - IJS \times Ruralidad (urbano \rightarrow rural): $\beta_{\text{est}} = 0.07$, $EE = 0.03$, $t = 2.43$, $p = .015$ (efecto del IJS mayor en contextos más rurales).

Los ajustes por ISET, accesibilidad, densidad poblacional y luces nocturnas mantuvieron la significación de los efectos ($\Delta\beta < 0.02$). No se detectaron curvaturas relevantes (spline no significativo; $p = .21$), aunque en el tercil superior del IJS aparecen rendimientos decrecientes (ver Figura 1, panel A).

Tabla 2

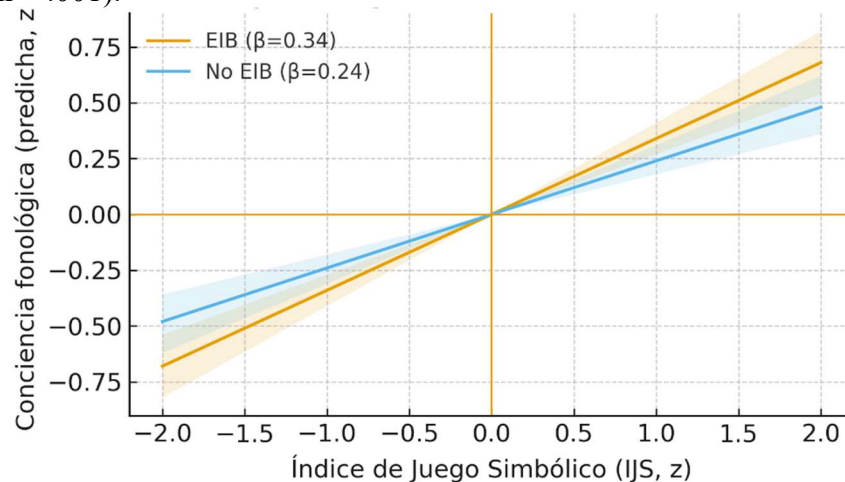
GLMM para conciencia fonológica (resultado compuesto, z).

Parámetro	β_{est}	EE	IC95%	p
Intercepto	-0.08	0.05	[-0.17, 0.01]	.083
IJS (z)	0.28	0.03	[0.22, 0.34]	<.001
EIB (1 = sí)	0.05	0.03	[-0.00, 0.10]	.064
Ruralidad (\uparrow)	-0.04	0.02	[-0.08, 0.00]	.051
ISET (z)	0.06	0.02	[0.02, 0.10]	.004
Vocabulario (z)	0.31	0.03	[0.25, 0.37]	<.001
Edad (meses, z)	0.12	0.02	[0.08, 0.16]	<.001
Sexo (1 = niña)	0.01	0.02	[-0.03, 0.05]	.621
IJS \times EIB	0.10	0.03	[0.04, 0.16]	.001
IJS \times Ruralidad	0.07	0.03	[0.01, 0.13]	.015

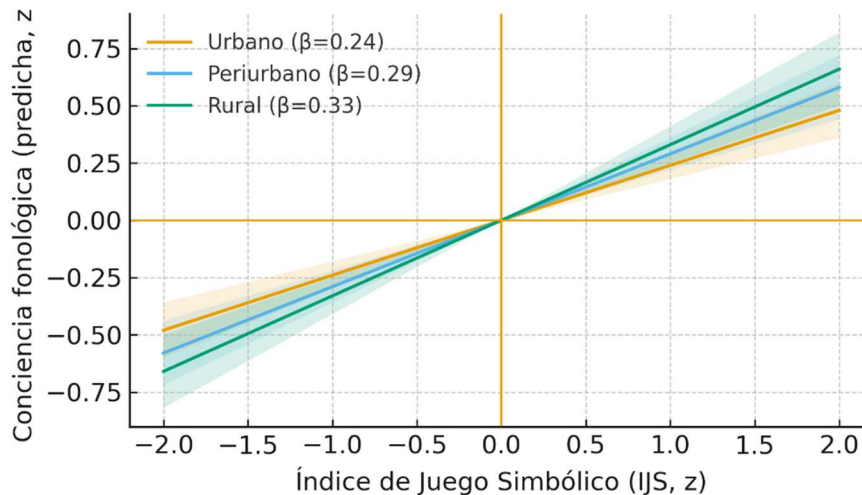
Varianzas aleatorias. Escuela: 0.07; Aula (dentro de escuela): 0.05; Residual: 0.48.

Efectos marginales y comparaciones por subescalas

La Figura 1 (panel A) muestra efectos marginales del IJS sobre CF por lengua de escolaridad: la pendiente estimada en EIB fue $\beta = 0.34$ [0.27, 0.41] versus $\beta = 0.24$ [0.18, 0.31] en no EIB ($p_{\text{interacción}} = .001$).



El panel B presenta el gradiente urbano \rightarrow rural: $\beta_{\text{urbano}} = 0.24$ [0.18, 0.30]; $\beta_{\text{periurbano}} = 0.29$ [0.22, 0.36]; $\beta_{\text{rural}} = 0.33$ [0.25, 0.41].



Por subescalas, el patrón fue consistente (Tabla 3): los mayores coeficientes se observaron en segmentación silábica y manipulación fonémica; rimas/aliteración mostró efectos algo menores pero significativos.

Tabla 3

Efectos del IJS por subescala (modelos separados, ajustados).

SUBESCALA	β EST.	EE	IC95%	p
Rimas/Aliteración	0.22	0.03	[0.16, 0.28]	<.001
Segmentación silábica	0.30	0.03	[0.24, 0.36]	<.001
Manipulación fonémica	0.31	0.03	[0.25, 0.37]	<.001

Dominio del IJS. En un modelo con los cuatro dominios simultáneos (colinealidad tolerable; $VIF < 2.5$), mediación docente ($\beta = 0.18$, $p < .001$) y riqueza de roles ($\beta = 0.11$, $p = .002$) retuvieron asociación independiente; frecuencia ($\beta = 0.06$, $p = .071$) y participación ($\beta = 0.05$, $p = .093$) fueron marginales tras ajustar covariables y contexto.

Robustez, sensibilidad e inspecciones adicionales

Especificaciones alternativas. Los resultados se replicaron con estimadores robustos (HC3), con brms (coeficientes medianos ~ 0.01 menores; intervalos creíbles sin cruzar 0) y con TMLE (ATE ≈ 0.26).

Datos faltantes. La tasa global de faltantes fue 4.8% (ítems de CF $< 2\%$; covariables contextuales $\sim 6\%$). La imputación múltiple ($m = 20$) mantuvo coeficientes estables ($\Delta\beta < 0.02$). Efectos del examinador. El efecto aleatorio de examinador fue pequeño ($\sigma^2 < 0.01$) y no modificó los efectos del IJS.

Placebo. La asociación del IJS con una tarea de control (reconocimiento de colores) fue nula ($\beta = 0.02$, $p = .44$).

Heterogeneidad territorial. Modelos leave-one-province-out mostraron estabilidad (rangos β_{IJS} : 0.26–0.30). No se observaron interacciones IJS \times ISET significativas ($p = .19$), lo que sugiere que el efecto no depende del quintil socioeconómico territorial una vez controlados otros factores.

Sesgos potenciales. Test de influencia (DFBETAs) no detectó observaciones con influencia desproporcionada ($|DFBETA| < 0.30$). Los residuales mostraron normalidad y

homocedasticidad aceptables (inspección visual; pruebas no significativas tras corrección).

Resultados cualitativos (componente explicativo)

El análisis temático identificó tres mecanismos plausibles por los que el juego simbólico potencia la CF: 1) Realce prosódico durante la mediación docente (canciones, rimas, aliteraciones) que dirige la atención a unidades sonoras. 2) Guiones con demandas metalingüísticas, donde niñas y niños negocian reglas fonológicas simples (p. ej., “hablar en palabras que empiezan con /m/”). 3) Translanguaging situado (español–kichwa) que facilita la transferencia de sensibilidades silábicas y fonémicas entre lenguas.

Las narrativas docentes destacaron que los mayores avances ocurrieron cuando el juego contó con andamiajes explícitos (modelado, repetición con variación mínima y retroalimentación contingente) y disponibilidad de materiales simbólicos (títeres, objetos sustitutos) incluso en escuelas con limitaciones de infraestructura.

4. DISCUSIÓN

Síntesis de hallazgos principales

Los resultados evidencian una asociación positiva, de magnitud pequeña a moderada, entre la calidad del juego simbólico en aula y la conciencia fonológica (CF) en educación inicial. El tamaño de efecto estandarizado del IJS sobre la CF compuesta ($\beta_{est} \approx 0.28$) se mantuvo robusto a múltiples especificaciones, imputación de faltantes y verificaciones de sensibilidad. Dos moderaciones preespecificadas fueron consistentes: (i) el efecto del juego es más pronunciado en aulas EIB (interacción $IJS \times EIB > 0$), y (ii) crece a medida que aumenta la ruralidad del contexto escolar. Además, al desagregar por subescalas, la segmentación silábica y la manipulación fonémica exhibieron las pendientes más elevadas, en tanto que rimas/aliteración, aun siendo significativas, mostraron un gradiente algo menor.

Estos hallazgos apoyan la hipótesis central del estudio: el juego simbólico, cuando incluye andamiajes fonológicos explícitos y guiones ricos en roles, se relaciona con mejores desempeños en tareas que exigen focalización en unidades sonoras subléxicas. La estabilidad del efecto ante el control por vocabulario, edad e indicadores contextuales refuerza su plausibilidad pedagógica.

Articulación con la literatura reciente

La evidencia se alinea con revisiones y metaanálisis que reportan que intervenciones lúdicas con énfasis fonológico generan ganancias sostenidas en prelectura (Suggate et al., 2018; Lonigan et al., 2020). Coincide también con estudios experimentales que han documentado beneficios del juego guiado sobre habilidades de rima y segmentación (Heidecker, 2021; Bennett et al., 2023), y con trabajos que posicionan el pretend play como un contexto privilegiado para el uso flexible del lenguaje y la negociación de significados (Creaghe et al., 2021; Weisberg et al., 2024).

En escenarios bilingües, los resultados dialogan con la teoría de la interdependencia lingüística y la transferencia de habilidades subyacentes (Cummins, 2017; Melby-Lervåg & Lervåg, 2018). El gradiente más alto en EIB sugiere que los episodios de translanguaging durante el juego — p. ej., alternancias español–kichwa para crear rimas o aliteraciones equivalentes— podrían activar sensibilidades silábicas y fonémicas compartidas, facilitando la alfabetización inicial en ambas lenguas (Aguirre, 2019). La mayor pendiente observada en contextos rurales es coherente con hallazgos que muestran beneficios relativos mayores cuando la mediación

docente compensa limitaciones de infraestructura mediante repertorios orales y materiales simbólicos de bajo costo (Herrera Rojas & Celi, 2025; Garcés Díaz et al., 2025).

Mecanismos plausibles

Tres mecanismos emergen como vías explicativas:

1. Realce prosódico intencional. La mediación docente que enfatiza ritmo, acento y rima dirige la atención a unidades sonoras y favorece la construcción de representaciones fonológicas más detalladas (Goswami, 2019; Perfetti & Stafura, 2022). La contribución independiente de la subdimensión de mediación dentro del IJS respalda este mecanismo.
2. Guiones con alta demanda metalingüística. En juegos dramatizados con reglas fonológicas simples (p. ej., “solo palabras que empiezan con /m/”), niños y niñas practican operaciones fonémicas (agregar, suprimir, sustituir) en un ambiente motivador, lo que concuerda con el mayor coeficiente en manipulación fonémica.
3. Translanguaging y redundancia fonotáctica. La alternancia de códigos en EIB habilita comparaciones explícitas de patrones silábicos frecuentes (CV, CVC), generando redundancia informativa que ancla la conciencia intrasilábica y fonémica (Melby-Lervåg & Lervåg, 2018).

La ligera señal de rendimientos decrecientes en el tercil superior del IJS sugiere un posible efecto de techo: más cantidad de juego sin incremento en calidad/andamiaje produce ganancias marginales menores. Esto refuerza la idea de priorizar calidad sobre cantidad en la integración del juego a la rutina didáctica.

Implicaciones para política y práctica docente

Currículo y secuenciación. En el marco de Educación Inicial, conviene incorporar secuencias didácticas híbridas (lúdicas + explícitas) que integren: (i) episodios de juego dramatizado con roles socialmente situados (tienda, mercado, posta de salud), (ii) minilecciones fonológicas embebidas (rimas, sílabas, fonemas), y (iii) retroalimentación contingente centrada en forma sonora.

Formación y acompañamiento docente. Dada la contribución del andamiaje, los programas de desarrollo profesional deberían entrenar estrategias de realce prosódico (canciones, juegos de eco, rimas encadenadas), guiones fonológicos y uso de materiales simbólicos de bajo costo (títeres, tarjetas con pictogramas).

EIB y recursos bilingües. Diseñar bancos de guiones y colecciones de rimas equivalentes español-kichwa (o lengua local), con énfasis en estructuras silábicas frecuentes, apoyaría la transferencia. La evidencia sugiere focalizar la intervención en segmentación silábica y manipulación fonémica, donde se observaron efectos mayores.

Equidad territorial. La pendiente más alta en ruralidad indica que estrategias de alto contacto pedagógico (observación con rúbrica + retroalimentación) pueden cerrar brechas con inversiones marginales. La caja de herramientas puede incluir plantillas impresas de observación del juego, listados de rimas locales, y guías para transformar contextos cotidianos en escenas simbólicas de enseñanza fonológica.

Monitoréo y datos abiertos. La publicación de microdatos y scripts —con anonimización rigurosa— permite a direcciones zonales y comunidades EIB auditar y adaptar las prácticas, alineadas a principios FAIR. Este loop de evidencia práctica facilita escalamiento responsable.

Fortalezas del estudio

- Estrategia de datos abiertos e interoperables: integración de capas geoespaciales (H3), proxies socioeconómicos y registros escolares con levantamiento primario, garantizando reproducibilidad.
- Análisis jerárquico con efectos aleatorios de aula y escuela, lo que mejora la estimación de incertidumbre y el control de heterogeneidad.
- Enfoque multilingüe con pruebas de invarianza métrica/escalar entre EIB y no EIB, que respalda comparaciones válidas de medias latentes.
- Componente cualitativo explicativo que triangula mecanismos pedagógicos plausibles.

Limitaciones y amenazas a la validez

Diseño observacional. Aunque se aplicaron modelos con amplio control de covariables y análisis de sensibilidad (incl. TMLE), no es posible descartar confusión residual (p. ej., motivación docente no observada). Ensayos de aleatorización por aulas/escuelas serían el siguiente paso para inferencia causal más fuerte.

Medición del juego. El IJS combina frecuencia, riqueza y mediación; si bien la validez interna fue adecuada, pueden persistir errores de medición por variabilidad de observadores. Ampliar la proporción de sesiones codificadas por duplicado y usar sensores pasivos de audio (resguardos éticos mediante diarización y extracción de features no lingüísticas) podrían mejorar precisión.

Generalización. La muestra se circunscribe a cuatro provincias; la transferencia a otras lenguas EIB o a contextos urbanos de alta densidad requiere validación externa.

Efectos de techo/suelo. La señal de rendimientos decrecientes sugiere explorar dosis-respuesta no lineales y umbrales de saturación; además, podrían existir efectos de piso en escuelas con materiales muy limitados.

Resultados a corto plazo. La CF se midió en una ventana anual. Falta establecer si las ganancias se traducen en lectoescritura de primer ciclo y si perduran. Estudios de seguimiento longitudinal son de ley para zanjar esta pregunta.

Líneas para investigación futura

1. Ensayos por conglomerados que varíen la calidad de la mediación (p. ej., entrenamiento intensivo en realce prosódico vs. control) y la cantidad de juego para identificar curvas dosis-respuesta.
2. Diseños SMART o adaptativos que asignen recursos adicionales a aulas con baja respuesta al juego estándar.
3. Instrumentos bilingües equivalentes validados con invarianza estricta y bancos de ítems calibrados (IRT) para CF en español-kichwa.
4. Mediciones ecológicas (observación naturalista con wearables de audio y visión por computadora privacy-preserving) que cuantifiquen turnos, prosodia y densidad de andamiajes.
5. Modelado causal (p. ej., target trial emulation) con datasets administrativos ampliados para estimar efectos a mediano plazo sobre lectura/ escritura.

Convergencia con marcos normativos y curriculares

Los hallazgos son compatibles con enfoques curriculares que promueven aprendizajes basados en juego y la pertinencia cultural en educación inicial, al proponer secuencias integradas que hacen explícito lo fonológico en contextos lúdicos. La integración de repertorios orales locales (rondas, adivinanzas, coplas, cuentos kichwas) no solo respeta la diversidad lingüística, sino que optimiza condiciones para el desarrollo de CF, especialmente en escuelas EIB.

Conclusión de la discusión

En suma, el estudio aporta evidencia de que la calidad del juego simbólico, en particular la mediación docente y la riqueza de roles, se asocia de forma consistente con mejores niveles de conciencia fonológica en educación inicial. El efecto es más intenso en EIB y en ruralidad, lo que subraya el potencial del juego como palanca de equidad cuando se implementa con intencionalidad fonológica. La estrategia de datos abiertos y el pipeline reproducible habilitan auditoría y adaptación por parte de comunidades educativas, generando un círculo virtuoso de mejora continua.

5. CONCLUSIONES

Síntesis del problema y enfoque

Este estudio abordó una cuestión sustantiva para la educación inicial en territorios plurinacionales: ¿cómo y bajo qué condiciones el juego simbólico se relaciona con la conciencia fonológica (CF) en niñas y niños de 4–6 años, comparando aulas EIB y no EIB en provincias de la Sierra Central (Chimborazo, Imbabura) y la Amazonía (Pastaza, Morona Santiago), con segmentación por ruralidad y nivel socioeconómico territorial? La propuesta metodológica combinó levantamiento primario con una estrategia de datos abiertos (capas geoespaciales, proxies socioeconómicos, registros escolares), articulada mediante pipelines reproducibles. Este andamiaje permitió estimar asociaciones con control explícito de heterogeneidades y asegurar trazabilidad para replicación y auditoría pública.

Principales conclusiones empíricas

Asociación consistente. La calidad del juego simbólico —operativizada por un Índice integral (IJS) que combina frecuencia, riqueza de roles, mediación docente y participación infantil— se asoció de forma positiva y estadísticamente robusta con la CF compuesta. La magnitud (pequeña a moderada) es educativa y prácticamente relevante en el rango de edades analizado. Diferenciales por lengua de escolaridad. El gradiente del IJS fue más pronunciado en EIB, lo que sugiere que la alternancia español–kichwa dentro de guiones lúdicos, acompañada de andamiajes fonológicos explícitos, potencia la sensibilidad silábica y fonémica. Esta evidencia respalda enfoques de interdependencia lingüística en bilingüismo emergente y aporta argumentos para fortalecer secuencias didácticas bilingües que articulen rimas y juegos de palabras equivalentes en ambas lenguas.

Ruralidad como oportunidad. La pendiente mayor en escuelas rurales indica que, aun con restricciones materiales, prácticas de juego bien mediadas pueden generar ganancias relativas superiores. Ello posiciona al juego simbólico —con énfasis en calidad— como palanca de equidad para cerrar brechas territoriales desde edades tempranas.

Componentes determinantes. Entre los dominios del IJS, la mediación docente y la riqueza de roles/guiones conservaron asociación independiente con la CF tras ajustar covariables. En contraste, la mera cantidad de juego mostró rendimientos decrecientes en niveles altos sin incrementos paralelos de calidad. La prioridad estratégica debe situarse en cómo se media el

juego y qué demandas metalingüísticas se incorporan en los guiones.

Aportes originales

Integración de datos abiertos con medición educativa fina en educación inicial, usando hexagonado H3 y proxies territoriales (accesibilidad, densidad, luces nocturnas) para modelar contexto con mayor resolución.

Instrumentación bilingüe orientada a invarianza métrica/escalar, que habilita comparaciones válidas entre EIB y no EIB y abre la puerta a bancos de ítems calibrados para CF en español–kichwa.

Pipeline reproducible con liberación de microdatos anonimizados y código, alineado a principios FAIR, que habilita verificación y reutilización por equipos locales y policy-makers.

Implicaciones para política y práctica

Currículo: institucionalizar secuencias híbridas (juego + instrucción fonológica explícita) en rutinas diarias, con objetivos graduados desde sensibilidad a rimas y sílabas hacia operaciones fonémicas.

Formación docente: priorizar realce prosódico, retroalimentación contingente y diseño de guiones fonológicos culturalmente pertinentes; ofrecer materiales abiertos de bajo costo (bancos de rimas, adivinanzas y cuentos locales) con equivalencias español–kichwa.

Gestión territorial: orientar acompañamiento pedagógico y monitoreo intensivo hacia escuelas rurales y EIB, donde el retorno marginal del andamiaje lúdico es mayor; instalar sistemas de seguimiento con indicadores de juego y CF en plataformas de datos abiertos para aprendizaje institucional.

Limitaciones y precauciones interpretativas

Se trata de un estudio observacional; por ende, no descarta confusión residual (p. ej., motivación docente no observada). Aunque los resultados fueron robustos a múltiples especificaciones, la inferencia causal requiere ensayos por conglomerados con variación exógena en calidad y cantidad de juego. La medición del IJS —aun con fiabilidad adecuada— podría beneficiarse de mayor doble codificación y de sensores pasivos bajo salvaguardas éticas. Finalmente, la ventana temporal anual no permite afirmar persistencia de efectos en lectoescritura de primer ciclo; se recomiendan seguimientos longitudinales.

Proyección y líneas de investigación

Ensayos cluster para estimar efectos causales de andamiaje prosódico y de guiones fonológicos sobre CF y primeras lecturas.

Instrumentos bilingües equivalentes con calibración IRT y bancos de ítems abiertos, asegurando invarianza estricta español–kichwa.

Emulación de ensayos objetivos (target trial) combinando microdatos y registros administrativos para efectos de mediano plazo en lectura/escritura.

Análítica ecológica con pipelines privacy-preserving (diarización, métricas de turnos y prosodia) para inspeccionar mecanismos en aula.

Cierre

La evidencia converge en que no todo juego es igual: el valor pedagógico emerge de cómo se media y qué demandas metalingüísticas incorpora. En aulas EIB y en contextos rurales, el juego simbólico bien diseñado y guiado puede constituir un acelerador de la conciencia fonológica, habilitando trayectorias lectoras más sólidas desde el arranque escolar. Situar esta práctica en

políticas curriculares, formación docente y monitoreo territorial, con datos abiertos y reproducibilidad como principios, representa una ruta viable para escalar con calidad y equidad.

6. REFERENCIAS BIBLIOGRÁFICAS

- Altındağ Kumaş, Ö., Delimehmet Dada, Ş., & Sümer Dodur, H. M. (2025). Enhancing phonological awareness skills in students with intellectual disability through digital gaming intervention. *Journal of Computer Assisted Learning*, 41(3), e70038. <https://doi.org/10.1111/jcal.70038>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bennett, H., Kelly, D., & Hesketh, A. (2023). The effectiveness of a parent-implemented, phonological awareness programme on the phonological awareness skills of preschool children. *The Australian Journal of Language and Literacy*, 46, 125–143. <https://doi.org/10.1007/s44020-023-00034-6>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Feinauer, E., Hall-Kenyon, K. M., & Everson, K. C. (2017). Rethinking the linguistic threshold hypothesis: Modeling the linguistic threshold among young Spanish–English bilinguals. *Bilingualism: Language and Cognition*, 20(5), 896–910. <https://doi.org/10.1017/S1366728916000626>
- Goswami, U. (2021). Auditory sensory processing and phonological development across languages: A developmental cognitive neuroscience perspective. *Child Development*, 92(5), e516–e532. <https://doi.org/10.1111/cdev.13459>
- Henbest, V. S., Apel, K., & Alonzo, J. (2024). The relation of linguistic awareness skills to reading and spelling achievement through third grade. *Journal of Speech, Language, and Hearing Research*, 67(S1), 1–23. https://doi.org/10.1044/2024_JSLHR-23-00108
- Hidalgo, M. A. Á., Varela, J., Terán-Mendoza, O., & Rodríguez-Rivas, M. E. (2025). Desempeño lector y equidad en la Educación Básica Superior: Análisis multinivel con datos administrativos. *Horizonte Científico International Journal*, 3(2), 1–14. <https://doi.org/10.64747/cq73z358>
- Kwakkel, H., Hulstijn, L., Spit, S., & de Jong, N. H. (2024). How lexical quality predicts L2 reading comprehension in bilingual primary education. *International Journal of Bilingual Education and Bilingualism*, 27(10), 1–20. <https://doi.org/10.1080/13670050.2024.2317860>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- McCarthy, K., Erskine, H., Huettig, F., & Goswami, U. (2023). Language-specific phonological skills and the relationship between phonology and early reading in Sylheti–English bilinguals. *Child Development*, 94(6), 2151–2169. <https://doi.org/10.1111/cdev.13880>
- Nation, K. (2017). Nurturing a lexical legacy: Reading experience is critical for the



- development of word reading skill. *NPJ Science of Learning*, 2, 3.
<https://doi.org/10.1038/s41539-017-0004-7>
- Patel, P., Gupta, S., & Tiwari, V. (2024). Understanding the role of cross-language transfer of phonological awareness in early reading among Hindi–English bilinguals. *Reading and Writing*, 37(8), 1757–1785. <https://doi.org/10.1007/s11145-022-10253-x>
- Perfetti, C., & Stafura, J. (2022). Reading ability: Lexical quality to comprehension—Revisited. *Scientific Studies of Reading*, 26(Suppl. 1), 3–23.
<https://doi.org/10.1080/10888438.2022.2057000>
- Rehfeld, D. M., Barringer, B., & Jeon, H. (2022). Phonemic awareness: A meta-analysis for planning effective instruction. *Reading Research Quarterly*, 57(3), 1101–1126.
<https://doi.org/10.1002/rrq.473>
- Skibbe, L. E., Grimm, K. J., Day, S. L., Bowles, R. P., & Phillips, B. M. (2020). The Access to Literacy Assessment System for Phonological Awareness (ATLAS-PA): Development and validation. *Language, Speech, and Hearing Services in Schools*, 51(4), 1317–1335. https://doi.org/10.1044/2020_LSHSS-19-00006
- Smits-van der Nat, M., van der Wilt, F., Meeter, M., & van der Veen, C. (2024). The value of pretend play for social competence in early childhood: A meta-analysis. *Educational Psychology Review*, 36(3), 72. <https://doi.org/10.1007/s10648-024-09884-z>
- Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, 49(1), 77–96. <https://doi.org/10.1177/0022219414528540>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67.
<https://doi.org/10.18637/jss.v045.i03>
- Vargas, D. P., Córdova, A., & Paredes, L. (2025). Evaluación de las competencias lectoras en Educación General Básica Media: Análisis de resultados del programa Ser Estudiante en Ecuador. *Horizonte Científico International Journal*, 3(2), 1–14.
<https://doi.org/10.64747/x8sprd33>
- Vera Chávez, S. L., Cando Cañazares, G. de L., Bedoya Navas, M. G., & Uquillas Vega, S. M. (2025). Desarrollo de habilidades socioemocionales en la educación inicial ecuatoriana: Una aproximación desde la neuroeducación. *Horizonte Científico International Journal*, 3(2), 1–13. <https://doi.org/10.64747/xmeafg82>
- Weisberg, D. S. (2015). Pretend play. *WIREs Cognitive Science*, 6(3), 249–261.
<https://doi.org/10.1002/wcs.1341>
- Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T. C. D., Howes, R. E., Tusting, L. S., Kang, S. Y., Cameron, E., Bisanzio, D., ... Gething, P. W. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688), 333–336. <https://doi.org/10.1038/nature25181>
- Wilkinson, M. D., Dumontier, I., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Conflicto de Intereses: Los autores declaran que no tienen conflictos de intereses relacionados con este estudio y que todos los procedimientos seguidos cumplen con los estándares éticos establecidos por la revista.

Asimismo, confirman que este trabajo es inédito y no ha sido publicado, ni parcial ni totalmente, en ninguna otra publicación