

Evaluación de la competencia intercultural de estudiantes de ELE: estudio cuantitativo de una intervención didáctica en un contexto universitario pluricultural

Cuadernos CANELA, 37, pp. 157-182

Recibido: 29-IX-2025

Aceptado: 17-XII-2025

Publicado, versión impresa: 1-V-2026

ISSN 1344-9109

Publicado, versión electrónica: 1-V-2026

ISSN 2189-9568

© Los autores 2026

canela.org.es

Jaime Recuero Díaz

Universidad de Hokkaido, Sapporo, Japón

Eduardo Vila López

Universidad de Valencia, Valencia, España

Resumen

Este artículo explora el impacto de una sesión intensiva de 100 minutos diseñada para potenciar la competencia intercultural de 60 estudiantes universitarios de ELE en Japón de nivel inicial. La intervención combinó tareas de producción escrita sobre rutinas culturales, actividades de contraste guiado y metarreflexión grupal, evaluadas mediante un cuestionario de ocho ítems que cubre conocimientos, conciencia cultural, tolerancia a la ambigüedad y flexibilidad conductual. Para valorar los cambios, se emplearon pruebas *t* para muestras relacionadas, complementadas con métodos no paramétricos, remuestreo *bootstrap* y modelos de efectos mixtos. Además, los ítems se agruparon en un índice compuesto basado en una versión adaptada del modelo INCA (Consejo de Europa, 2004), lo que permitió analizar tanto los avances puntuales como el progreso global. Los resultados revelan mejoras entre el pretest y el postest, especialmente en las habilidades procedimentales, y aportan evidencia a favor de microsecuencias didácticas intensivas para desarrollar la competencia intercultural. Se proponen estudios con grupos control, seguimientos longitudinales y metodologías mixtas para evaluar la sostenibilidad de los efectos y reforzar la validación psicométrica del instrumento.

Palabras clave

Competencia intercultural, evaluación formativa, análisis cuantitativo, integración de contenido y lengua, español como lengua extranjera (ELE).

Introducción

La competencia intercultural ha cobrado protagonismo en la enseñanza de lenguas como respuesta al creciente encuentro entre hablantes de diversas procedencias. Aunque muchos estudios sobre la interculturalidad han abordado ampliamente cuestiones teóricas y experiencias cualitativas, son todavía limitadas las investigaciones cuantitativas que midan cambios en actitudes interculturales tras microintervenciones aplicadas en niveles elementales. Esta carencia resulta especialmente relevante en contextos universitarios con calendarios

académicos restringidos, donde es necesario demostrar la viabilidad de acciones pedagógicas breves y replicables.

El presente estudio examina el impacto de una intervención didáctica de 100 minutos en el desarrollo inicial de la conciencia intercultural en 60 estudiantes de ELE de nivel A1 de una universidad del sur de Japón caracterizada por una alta proporción de estudiantes internacionales.

El objetivo principal de la investigación es evaluar cambios pretest y postest en tres dimensiones consideradas clave para fases iniciales del aprendizaje intercultural, tomando como marco operativo el modelo INCA (Consejo de Europa, 2004): conciencia cultural (el interés y la valoración de las prácticas culturales ajenas), tolerancia a la ambigüedad (la capacidad de manejar situaciones donde las normas difieren de las propias) y flexibilidad conductual (la habilidad de adaptación a normas y prácticas culturales distintas). Para ello, se empleó un diseño pretest y postest con un cuestionario en escala Likert y una sesión en formato de trabajo grupal guiado centrada en rutinas y escenarios de choque cultural.

La contribución del trabajo es doble: aportar evidencia empírica sobre la eficacia potencial de microintervenciones interculturales en A1 y ofrecer un protocolo práctico y replicable para docentes que deseen integrar objetivos interculturales en cursos de lenguas extranjeras. En la sección metodológica se exponen con detalle los instrumentos, el proceso de validación y las decisiones analíticas. Los resultados permiten discutir implicaciones didácticas y futuras vías de investigación y sugieren una tendencia hacia el fortalecimiento de actitudes de apertura, curiosidad y disposición al contacto intercultural tras una intervención didáctica en el aula que incluye una reflexión guiada. Estos hallazgos, presentados con visualizaciones descriptivas, apoyan la viabilidad de integrar objetivos interculturales en niveles elementales y en contextos donde resulta necesario optimizar el tiempo disponible en el aula.

1. Marco teórico

La enseñanza de lenguas extranjeras ha experimentado una transformación paradigmática en las últimas décadas, transitando desde enfoques centrados exclusivamente en la competencia lingüística hacia modelos integrales que incorporan la dimensión intercultural como eje formativo esencial (Byram, 1997; Instituto Cervantes, 2006). Este cambio, impulsado por organismos internacionales como el Consejo de Europa (2020) y la UNESCO (2017), responde a las demandas de sociedades globalizadas donde la comunicación trasciende fronteras y requiere habilidades para negociar significados en contextos multiculturales. Como señala el *Plan Curricular del Instituto Cervantes* (2006), la interculturalidad no es un «contenido añadido», sino un componente integrado en una triple dimensión: cultura (referentes factuales), sociocultura (saberes y comportamientos) y competencia intercultural (habilidades y actitudes para comprender e integrar fundamentos culturales compartidos). Según el *Marco Común Europeo de Referencia*, la competencia intercultural implica no solo conocimientos objetivos, sino habilidades para «mediar entre sistemas culturales» a partir de una conciencia reflexiva que examina críticamente las propias referencias culturales (Consejo de Europa, 2020, p. 103).

Bajo esta perspectiva, el aula de lenguas se redefine como un espacio dialógico y crítico (Kramsch, 1993; Guilherme, 2002) donde se fomenta la formación de aprendices capaces de actuar localmente con una conciencia global, es decir, agentes que median entre culturas, gestionan ambigüedades y cuestionan estereotipos (Liddicoat y Scarino, 2013; Byram,

1997). Este enfoque supera la mera transmisión de estructuras gramaticales para priorizar la negociación de identidades culturales y la deconstrucción de jerarquías presentes en los materiales didácticos (González Plasencia, 2019). Además, la noción de hablante intercultural (Byram, 1997; Consejo de Europa, 2020) sustituye al ideal del hablante nativo, centrándose en la capacidad de mediación y en la responsabilidad ética del aprendiente. Algunos autores que investigan la educación crítica y la alfabetización cultural (Porto, 2014; Porto, 2016; Porto y Byram, 2015a) vinculan además la competencia intercultural con la justicia social y señalan la importancia de reconceptualizar el aula como un espacio de cuestionamiento de jerarquías culturales.

Sin embargo, como advierte González Plasencia (2019), la integración efectiva de esta competencia enfrenta tensiones teóricas y prácticas. La marginación histórica de la antropología en la didáctica de lenguas ha causado que en los materiales educativos se perpetúen las representaciones esencialistas, ejemplificadas en estereotipos como la idea de que todos los españoles duermen la siesta. Ciertos autores como Sercu (2010) subrayan la ausencia de instrumentos holísticos que capten todas las dimensiones de la competencia intercultural. Por otro lado, Deardorff (2016) recuerda la complejidad de medir los procesos metacognitivos y de mediación en contextos educativos. En la misma línea crítica, Borghetti (2017) plantea dudas sobre la pertinencia metodológica de evaluar la competencia intercultural a través de instrumentos estandarizados, pues señala que estos tienden a simplificar procesos interactivos y situados como la negociación de significado o la mediación. Según este autor, de ahí surge la necesidad de complementar las mediciones cuantitativas con enfoques cualitativos y reflexivos para captar el carácter dinámico de la competencia intercultural. A pesar de estos retos, diversos autores han subrayado que lengua y cultura son indisociables (Fantini, 1995) y que la formación intercultural es inherente al proceso de aprendizaje lingüístico (Liddicoat y Scarino, 2013).

A partir de esta fundamentación teórica, es necesario explorar cómo se traduce la enseñanza de conciencia intercultural en contextos concretos de aula, especialmente en niveles iniciales. En estos escenarios, las limitaciones formales derivadas del tiempo lectivo reducido, la secuenciación curricular y el nivel inicial del alumnado no invalidan el desarrollo intercultural, sino que demandan estrategias pedagógicas específicas y bien andamiadas (Liddicoat y Scarino, 2013). Los estudios con aprendientes de nivel A1 muestran que actividades basadas en experiencias cotidianas —como el análisis de rutinas de alimentación o protocolos de saludo— pueden activar procesos de conciencia crítica cuando se acompañan de andamiajes y reflexiones guiadas (Porto, 2016). Estos hallazgos coinciden con las observaciones de Valls Campà (2023) en aulas japonesas, quien describe los desafíos de implementar la ciudadanía intercultural en cursos de ELE, y destaca la necesidad de que el profesorado tenga en cuenta las experiencias educativas previas de los estudiantes y la selección de unos contenidos significativos para promover ciertas prácticas reflexivas en contextos donde la diversidad cultural directa sea limitada. En consecuencia, y tal como advierten Deardorff (2006, 2016) y Liddicoat y Scarino (2013), la ausencia de enfoques etnográficos y reflexivos en la formación del profesorado limita la posibilidad de generar experiencias de aprendizaje verdaderamente interculturales, más allá de lo puramente descriptivo. En esta línea, diversos estudios han abordado la evaluación de la competencia intercultural desde metodologías cualitativas y etnográficas, y han recurrido al análisis de interacciones en el aula, diarios reflexivos, estudios de caso y producciones discursivas del alumnado para captar procesos de interpretación cultural y negociación de significado (Byram, 1997; García-Balsas, 2021).

Pese al consenso teórico en torno a la centralidad de la mediación cultural y la formación de aprendices críticos, diversos estudios han señalado que persisten importantes desafíos para llevar a la práctica estos enfoques. Entre ellos destacan la persistencia de representaciones esencialistas en materiales didácticos, las dificultades para evaluar procesos metacognitivos complejos y las lagunas en la formación del profesorado. En este contexto, resulta imprescindible traducir estos planteamientos teóricos en secuencias didácticas cuidadosamente estructuradas y en instrumentos de evaluación que atiendan la realidad de aulas concretas. Esta necesidad justifica el enfoque de la presente investigación, que busca explorar cómo pueden desplegarse procesos de conciencia intercultural en contextos elementales mediante una intervención didáctica intensiva y el uso de andamiajes específicos.

1.1. La evaluación de la competencia intercultural: tensiones teóricas y propuestas emergentes

La medición de la competencia intercultural plantea desafíos epistemológicos que van más allá de una simple traducción a indicadores medibles. Como advierte Deardorff (2016), reducir esta dimensión a escalas cuantitativas desatiende su naturaleza dinámica y contextual. Desde distintas propuestas teóricas se han priorizado aspectos diferentes del constructo: mientras que algunos modelos como el de Byram (1997) subrayan elementos observables —actitudes y habilidades de mediación—, los enfoques críticos como los de Guilherme (2002) sostienen que la auténtica comprensión intercultural se expresa en la habilidad para dismantelar mecanismos de poder presentes en discursos culturales, un aspecto que es difícil capturar con herramientas estandarizadas.

El proyecto INCA (Consejo de Europa, 2004) constituye un referente paradigmático en los intentos por operacionalizar la evaluación de la competencia intercultural a través de un enfoque multidimensional. Su marco describe seis competencias nucleares: tolerancia a la ambigüedad, flexibilidad conductual, conciencia comunicativa, descubrimiento cultural, respeto al otro y empatía. A su vez, estas están organizadas en tres niveles de dominio (básico, intermedio y completo). Lo innovador de la propuesta radica en su concepción procesual y longitudinal, que entiende el desarrollo intercultural como una progresión desde la simple identificación de diferencias culturales hasta la capacidad de gestionar interacciones complejas y conflictos interculturales de manera constructiva.

Junto a estos modelos centrados en la evaluación, otros enfoques han abordado la competencia intercultural como un proceso de desarrollo progresivo a lo largo del tiempo. El Modelo de Desarrollo de Sensibilidad Intercultural (DMIS; Bennett y Hammer, 2017) parte de la idea de que las personas no se relacionan con la diferencia cultural de manera homogénea, sino que experimentan una evolución en su manera de percibirla y gestionarla. Esta propuesta distingue entre fases caracterizadas por la negación o minimización de la diferencia y etapas posteriores de aceptación, adaptación e integración. En estos niveles más avanzados, el sujeto conseguiría desarrollar la capacidad de ajustar su comportamiento comunicativo y negociar significados en contextos culturalmente diversos. Una aportación relevante del DMIS es que permite diferenciar entre cambios externos observables —por ejemplo, en la manera de interactuar o de adaptar la conducta en situaciones interculturales— y transformaciones internas más profundas, vinculadas a actitudes, valores y marcos interpretativos, que requieren tiempos de exposición más prolongados. Desde esta perspectiva, las intervencio-

nes educativas breves pueden entenderse no como mecanismos de transformación completa, sino como experiencias que pueden activar unos cambios iniciales en la forma de percibir y afrontar la diferencia cultural, siempre que se apliquen en procesos formativos más amplios que favorezcan la reflexión continuada.

Ahora bien, diversos autores advierten que evaluar la interculturalidad implica necesariamente tomar decisiones sobre qué prácticas o actitudes se consideran legítimas, con el consiguiente riesgo de reproducir criterios etnocéntricos o jerárquicos entre culturas (Deardorff, 2009; Liddicoat y Scarino, 2013). También dificulta la evaluación de la competencia intercultural la falta de preparación específica del profesorado para valorar dimensiones actitudinales y cognitivas, lo que pone de relieve la necesidad de integrar la evaluación crítica de la interculturalidad en los programas de formación docente (Iglesias-Casal, 2021). Ante estas tensiones, diversos trabajos han puesto el acento en los procedimientos de evaluación de orientación cualitativa, que buscan captar la dimensión procesual de la competencia intercultural. Entre ellas destacan los portafolios reflexivos, el análisis de narrativas autobiográficas (Sercu, 2010) o las simulaciones de mediación (Deardorff, 2016). Sin embargo, autores como Sercu (2010) señalan la complejidad intrínseca de evaluar los planos procedimental y afectivo, pues al intentar medir actitudes como la empatía o la apertura ante la diversidad se correría el riesgo de juzgar la personalidad del alumno o de imponerle cómo debe sentir y reaccionar. Además, la interpretación de evidencias puede variar según la cultura del evaluador y del estudiante, lo que Landone (2013) identifica como un factor que distorsiona los resultados. Para enfrentar estas dificultades, Deardorff (2016) propone un enfoque multimétodo y multiperspectiva, centrado en el proceso de desarrollo de la competencia intercultural más que en la obtención de resultados finales.

En la misma línea, algunos estudios recientes aportan perspectivas valiosas sobre instrumentos de evaluación. Deardorff (2020), por ejemplo, argumenta que la triangulación de evidencias —a través de la autoevaluación, observación de interacciones y análisis de producciones escritas— es fundamental para obtener una imagen más precisa de la competencia y mitigar los sesgos culturales inherentes a cualquier proceso evaluativo. Estas ideas coinciden con la propuesta de Liddicoat y Scarino (2013) de rúbricas centradas en indicadores de la competencia intercultural que evalúan la capacidad de relativizar la cultura propia o negociar significados en la ambigüedad. En conjunto, estas perspectivas evidencian que evaluar la conciencia intercultural no es solo un desafío técnico, sino un proceso ético y pedagógico que requiere de instrumentos flexibles, reflexivos y culturalmente sensibles.

En los últimos años, muchas propuestas en evaluación han tendido a alejarse de una visión puramente técnica o instrumental para situarse en una comprensión más ética del aprendizaje. Desde esta perspectiva, evaluar no consiste únicamente en medir resultados, sino en generar espacios en los que el alumnado pueda reflexionar sobre su propio proceso y se le involucre en el diseño mismo de los criterios. En este sentido, diversos estudios sobre intercambios virtuales muestran que el uso de rúbricas compartidas y herramientas de autoevaluación favorece los procesos de reflexión crítica y diálogo intercultural en proyectos colaborativos (Mont y Masats, 2018; Dooly y O’dowd, 2018). De manera similar, las investigaciones en el ámbito universitario señalan que implicar a los estudiantes en la coconstrucción de criterios de evaluación contribuye a una mayor claridad, autonomía y conciencia del propio aprendizaje (Fraile *et al.*, 2017; Panadero y Jonsson, 2013). En paralelo, el uso de portafolios digitales reflexivos se ha consolidado como una vía especialmente adecuada para

documentar procesos de toma de conciencia cultural y promover una alfabetización digital crítica en contextos educativos mediados por tecnologías (Bilki, *et al.*, 2023; Hauck, 2019). En conjunto, estas aproximaciones parten de una concepción de la competencia intercultural como un proceso en continua evolución que difícilmente puede capturarse mediante indicadores estandarizados o mediciones aisladas. Así, evaluar implica asumir una posición pedagógica que reconoce la dimensión contextual del aprendizaje y sitúa al alumnado como agente activo en la elaboración del sentido intercultural (Guilherme, 2002).

1.2. Hacia un enfoque crítico de los materiales en la enseñanza intercultural

La crítica a los materiales didácticos tradicionales constituye una constante en la pedagogía intercultural contemporánea, especialmente cuando, más allá de su adscripción metodológica, estos se apoyan en concepciones de la cultura y del aprendizaje que tienden a simplificar la complejidad de los procesos interculturales. González Plasencia (2019) identifica un patrón recurrente en los manuales de ELE: la reducción de realidades culturales a esencialismos estáticos que presentan culturas como entidades homogéneas y descontextualizadas, perpetuando visiones simplificadas del tipo «los españoles son extrovertidos» o «los japoneses prefieren estar en silencio». Esa tendencia a la homogeneización también funciona como un mecanismo de poder que oculta las dinámicas de jerarquía y desigualdad, una problemática analizada críticamente en la literatura sobre estereotipos culturales en la enseñanza del inglés como lengua extranjera (Kumaravadivelu, 2003).

Frente a este panorama, emergen enfoques que replantean los recursos de enseñanza desde una perspectiva crítica. Algunas propuestas de ciudadanía intercultural, como las desarrolladas por Porto (2014; 2016), fomentan el análisis de producciones culturales auténticas —desde vídeos y blogs hasta testimonios y producciones colaborativas— como herramientas para desestabilizar narrativas hegemónicas, cuestionar identidades y promover la acción comunitaria. En este sentido, el trabajo de Porto (2016) muestra cómo el uso didáctico de este tipo de recursos en el aula de inglés favorece que el alumnado analice críticamente las relaciones de poder presentes en prácticas culturales cotidianas. Lo innovador de estas propuestas radica en desplazar el foco desde una descripción de «qué hacen los otros» hacia el examen crítico de los conflictos identitarios y las relaciones de poder que esas prácticas ponen de manifiesto. Esta orientación conecta con el objetivo fundamental de formar aprendices críticos y políticamente informados (Guilherme, 2002; Liddicoat y Scarino, 2013; Porto y Byram, 2015b), y ha sido desarrollada en investigaciones recientes en el ámbito del ELE que analizan cómo los aprendientes emplean recursos lingüísticos para construir posicionamientos críticos en contextos interculturales reales (García-Balsas, 2021).

Sin embargo, a pesar del valor de estos enfoques, la realidad de las instituciones a menudo impone limitaciones que dificultan aplicar estos enfoques críticos en determinados contextos educativos. En el ámbito universitario japonés, Valls Campà (2023) señala que la dependencia de manuales comerciales y la rigidez curricular pueden resultar en obstáculos para introducir propuestas críticas de competencia intercultural para la ciudadanía. A ello se suma la lógica de mercado de las editoriales, que tiende a favorecer la producción de materiales con representaciones culturales simplificadas con el objetivo de facilitar su uso inmediato en clase (Gray, 2010; Risager, 2006), lo que restringe la integración de miradas críticas. Ante este escenario, Liddicoat y Scarino (2013) proponen que el trabajo intercultural

ral en la enseñanza de lenguas se apoye en andamiajes pedagógicos que permitan incorporar una dimensión crítica en las prácticas habituales del aula, sin necesidad de sustituir completamente los materiales existentes.

Como señala González Plasencia (2019), es fundamental asumir que, a priori, ningún material didáctico es neutral, pues cada recurso encarna una determinada visión del mundo y puede contribuir tanto a reproducir como a cuestionar jerarquías culturales. A este respecto, las investigaciones desarrolladas en el marco del grupo CRITERION-ELE han mostrado cómo los materiales y las prácticas habituales en el aula de ELE incorporan posicionamientos ideológicos que requieren ser analizados críticamente (véase, por ejemplo, Cerezo y Bertolín, 2020; García-Balsas, 2021). En consecuencia, la función del docente no debe limitarse a la selección de contenidos, sino que debe fomentar una mirada crítica en el aula, de forma que le permita al alumnado identificar las estructuras de poder presentes en gestos cotidianos aparentemente simples —como un saludo, un horario o una costumbre— y convertir el espacio educativo en un lugar donde dichas prácticas puedan ser cuestionadas, y no solo reproducidas.

2. Metodología

En este apartado se describen los elementos metodológicos de este estudio. En primer lugar, se presenta el objetivo general de la investigación y el diseño adoptado. A continuación, se detalla la intervención didáctica y su secuencia de actividades, así como el perfil de los participantes. Seguidamente, se explican el instrumento de recogida de datos y los criterios del agrupamiento de los ítems. Por último, se exponen los procedimientos del análisis estadístico empleado para medir los cambios en el pretest y postest.

2.1. Objetivo

La finalidad de este estudio es evaluar si una intervención didáctica de 100 minutos, diseñada para desarrollar la competencia intercultural en estudiantes de ELE, produce cambios medibles en dimensiones cognitivas, actitudinales y procedimentales. El diseño fue de grupo único pretest y postest: se administró la misma encuesta de 8 ítems (escala Likert 1-6) al inicio del curso académico (medición previa, en adelante *pre*) y justo después de la sesión específica de intervención de 100 minutos (medición posterior, en adelante *post*), lo que permitió comparar las puntuaciones medias del mismo grupo. Durante el intervalo entre pre y post, el curso se desarrolló siguiendo el manual *Aula Internacional I Plus* (Difusión), el cual no contiene apartados explícitamente orientados al desarrollo de la competencia intercultural. Por ello, aunque el intervalo temporal entre el pretest y el postest no permite descartar completamente la influencia de otros factores contextuales, consideramos poco probable que el uso de este manual haya influido de manera directa en los resultados observados, dado que no incorpora contenidos ni actividades explícitamente orientados al desarrollo de la competencia intercultural.

La muestra final estuvo compuesta por 60 estudiantes de ELE que estudian español como lengua optativa en una universidad internacional al sur de Japón, con un nivel inicial de competencia lingüística (A1) y distribuidos en tres clases del mismo nivel impartidas en días y horarios diferentes. Para garantizar la comparabilidad, solo se incluyeron los cuestionarios completos en ambas mediciones y se descartaron las respuestas de participantes que solo

respondieron únicamente en una de las dos fases. En los 60 cuestionarios retenidos no se registraron datos faltantes, de modo que los análisis se realizaron con casos completos, sin necesidad de exclusión adicional ni imputación.

2.2. Diseño de la intervención

Desde el punto de vista didáctico, el diseño de la intervención se apoya en el planteamiento de Byram (1997), que subraya la necesidad de promover actividades de contraste cultural y reflexión crítica para el desarrollo de la competencia intercultural. Esta orientación es coherente, además, con enfoques integrados de contenido y lengua como el CLIL (Coyle *et al.*, 2010), que enfatizan la articulación entre el uso lingüístico, contenido y reflexión.

El diseño temporal del estudio se estructuró en tres fases: la medición previa (pre), la propuesta didáctica y la medición posterior (post). El primer día de clase del curso, con una duración total de 14 semanas, se administró la encuesta previa (pre) en una versión multilingüe (español, inglés y japonés), adaptada al nivel inicial del alumnado. Tras nueve semanas de docencia —equivalentes a 36 sesiones regulares de 100 minutos—, en una fase intermedia del curso coincidente con el trabajo sobre rutinas y horarios (gramática y léxico de rutinas, expresiones de tiempo y frecuencia), se llevó a cabo la intervención intensiva de 100 minutos. Al término de la sesión se aplicó la medición posterior (post), lo que permitió emparejar respuestas y comparar cambios a corto plazo asociados a la actividad. De forma práctica, se combinaron tareas individuales, comparativas y grupales orientadas a fomentar no solo el conocimiento sobre prácticas culturales, sino también la reflexión crítica sobre sus causas y consecuencias. El objetivo fue activar conocimientos, actitudes y habilidades en un formato intensivo, entendido como una intervención concentrada en una única sesión que integra actividades encadenadas de análisis, interacción y reflexión. La secuencia se describe a continuación:

1. Producción escrita individual. Se pidió al alumnado que preparara un breve texto descriptivo de carácter personal, en el que recogiera, mediante frases simples, sus rutinas cotidianas (por ejemplo, actividades y horarios habituales entre semana y durante el fin de semana), aprovechando el vocabulario y la gramática trabajados previamente. Esta tarea favorece la reactivación de conocimientos y provee material auténtico para la comparación intercultural.
2. Comparación guiada en gran grupo. En la sesión didáctica se pusieron en común esos textos para detectar diferencias y similitudes entre las rutinas de los distintos contextos socioculturales de procedencia del alumnado. La comparación se realizó con preguntas mediadoras del profesorado que orientaron la observación hacia fenómenos culturales relevantes, como, por ejemplo, horarios de comidas, prácticas sociales y normas implícitas.
3. Encuesta individual sobre rutinas. A continuación, los estudiantes completaron una breve encuesta contextualizada en su propia rutina (por ejemplo, «¿A qué hora almuerzan normalmente en tu país?»), para cuantificar percepciones y crear una línea base de respuestas personales. Aunque la formulación alude al país de referencia, las respuestas se interpretaron como aproximaciones situadas, vinculadas a los contextos socioculturales próximos del alumnado (familia, entorno social o experiencias previas), y no como representaciones homogéneas de una cultura nacional.

4. Trabajo en grupos pequeños (3 o 4 estudiantes por grupo). Tras compartir resultados individuales, el alumnado trabajó en pequeños grupos (*cf.* 2.3) con el fin de responder colectivamente a preguntas de reflexión más profundas sobre diferencias y similitudes culturales. En esta fase se les solicitó justificar sus respuestas y explorar causas: por ejemplo, «¿Qué choque cultural has tenido en Japón o en el extranjero y por qué? ¿Cómo podrías evitar o mitigar esos conflictos?». La dinámica de grupo buscó promover el diálogo, el contraste argumentado y la negociación de significados.
5. Metarreflexión y puesta en común. El profesorado orientó una discusión final en la que se invitó al alumnado a pasar del enunciado de diferencias superficiales a una reflexión sobre razones culturales subyacentes (estructuras sociales, hábitos, valores), promoviendo así la comprensión crítica.
6. Encuesta post. La intervención concluyó con la aplicación de la misma encuesta utilizada como pretest, lo que permitió emparejar respuestas y medir cambios inmediatos atribuibles a la experiencia.

Dado que el nivel de competencia en español del alumnado era todavía limitado, y para asegurar la eficacia comunicativa en las actividades grupales y la calidad de la información recogida, se permitió el uso de inglés como lengua vehicular en el trabajo en equipo y se aconsejó recurrir al español y japonés cuando fuera necesario. La encuesta también se presentó en tres versiones (español, inglés y japonés) para facilitar la comprensión y la fiabilidad de las respuestas en un contexto plurilingüe. Esta decisión se fundamenta en el marco del *translanguaging* (García y Li, 2014) y, de forma más específica, en el *pedagogical translanguaging* o translingüismo pedagógico (Cenoz y Gorter, 2022), que promueve el uso estratégico de las lenguas del alumnado para facilitar la interacción, la construcción de significado compartido y el desarrollo de la conciencia metalingüística. De este modo, se priorizó la interacción significativa en el trabajo en grupo y la recogida de datos válidos en la encuesta.

2.3. Perfil de los participantes

La investigación se realizó en una universidad de carácter internacional situada en el sur de Japón, con un alumnado diverso culturalmente y una fuerte presencia de estudiantes extranjeros. Los 60 estudiantes pertenecían a tres grupos distintos del mismo nivel (A1). Las clases se impartieron en días y horarios diferentes, pero siguieron un plan docente idéntico y participaron en la intervención de forma independiente. En cuanto a la composición de dichos grupos, los 60 participantes de esta pesquisa provenían de 14 nacionalidades distintas, de las cuales las más numerosas eran la japonesa (25 estudiantes), la indonesia (8), la birmana (6), la tailandesa y la china (4 ambos países).

2.4. Instrumento

Se empleó un cuestionario breve de ocho ítems. Si bien muchas escalas de competencia intercultural emplean formatos con número impar de categorías para incluir una opción neutral (5 o 7 puntos), en este estudio se optó por una escala de 6 puntos. Esta decisión fuerza al alumnado a posicionarse en una dirección, lo cual puede resultar ventajoso en contextos de formación breve donde interesa detectar cambios direccionales claros. El ítem P1 recoge conocimiento declarado sobre costumbres y tradiciones tras varios meses de curso; los ítems

P2-P8 evalúan aspectos vinculados con la competencia intercultural (valoración del aprendizaje cultural, tolerancia, interés, comodidad en la interacción, disposición a participar e identificación de similitudes sin juicios de valor). El texto íntegro del cuestionario y sus versiones en español, inglés y japonés (revisadas por hablantes nativos para asegurar equivalencia conceptual) figuran en el apéndice 1. La administración, que fue emparejada (pretest y postest), se obtuvo con el consentimiento informado de los participantes y se siguieron las normas éticas de la institución.

2.5. Agrupamiento de ítems y justificación

Se definieron dos niveles de análisis con fines interpretativos: (a) análisis ítem a ítem (P1-P8) y (b) un índice compuesto (P2-P8) para estimar el efecto inmediato de la sesión sin el sesgo del ítem acumulativo de la P1. Además, vinculamos los ítems con una versión operativa del modelo INCA (Intercultural Competence Assessment), que proporciona un marco explícito para evaluar dimensiones cognitivas, actitudinales y procedimentales de la competencia intercultural (Consejo de Europa, 2004; Prechtel y Davidson-Lund, 2007). Con este propósito, empleamos el siguiente mapeo: conciencia cultural = P4, P6 (valoración e interés por conocer costumbres y tradiciones); tolerancia a la ambigüedad = P2, P5, P8 (comodidad ante la diferencia y capacidad de observar sin juzgar); flexibilidad conductual = P3, P7 (habilidad y disposición para adaptar la conducta y participar en interacciones interculturales). La P1, que mide el conocimiento declarado acumulado tras más de 60 horas de clase, se analiza por separado para evitar que su carácter longitudinal sobredimensione la estimación del efecto puntual de la intervención de 100 minutos.

2.6. Análisis estadístico

En este apartado resumimos los procedimientos empleados para cuantificar los cambios pre-post tras la intervención ($n = 60$). Estos cambios se interpretan principalmente en relación con la intervención didáctica de 100 minutos, diseñada para activar procesos específicos de reflexión y actuación intercultural, sin perder de vista que otros factores vinculados al desarrollo general del curso pudieron también influir en los resultados. El objetivo fue (a) describir la evolución media de las respuestas, (b) evaluar si los cambios superan la variabilidad aleatoria y (c) estimar su magnitud práctica. Para ello, se siguió el siguiente protocolo.

En primer lugar, para describir la evolución pre- y posactividad, calculamos las medias y desviaciones estándar para cada uno de los ocho ítems (P1-P8) y para el índice compuesto (P2-P8). A continuación, empleamos una prueba t pareada para comparar las medias antes y después de la intervención didáctica intensiva de 100 minutos. El umbral de significación adoptado fue $\alpha = 0,05$. Para comprobar la validez, se evaluó la normalidad de las diferencias pretest y postest mediante el test de Shapiro-Wilk.

Para complementar la significación estadística, se calculó el d de Cohen emparejado¹ para medir la relevancia práctica de los cambios (Cohen, 1988) y obtuvimos su intervalo de confianza al 95 % mediante el remuestreo *bootstrap* con 5000 réplicas, siguiendo el procedimiento descrito por Efron y Tibshirani (1993). Dado el número de comparaciones realizadas (por ítem y por dimensión), se muestran en las tablas valores p ajustados y sin ajustar² (Benjamini y Hochberg, 1995; Bonferroni, 1936). Además, se realizaron comprobaciones

de robustez (pruebas no paramétricas, *bootstrap* y modelos mixtos) para garantizar que los resultados no dependieran de supuestos puntuales de la prueba *t*.

Finalmente, evaluamos la fiabilidad de la escala compuesta (P2-P8) mediante el alfa de Cronbach (α) y el coeficiente omega (ω)³, e inspeccionamos las correlaciones ítem-ítem para detectar posibles inconsistencias en las subescalas de dos ítems. Todos los análisis se llevaron a cabo en el lenguaje de programación R, con paquetes específicos para cada prueba. Véase el apéndice 2 para definiciones y referencias completas.

3. Análisis de resultados

En esta sección mostramos, de forma ordenada y concisa, la evidencia empírica sobre los cambios observados tras la intervención ($n = 60$). Primero, presentamos los resultados ítem por ítem (P1-P8) mediante la figura 1 y la tabla 1; a continuación, el índice compuesto P2-P8 (tabla 2) y, finalmente, el análisis por dimensiones apoyándonos en el modelo INCA (figura 2 y tabla 3). Las tablas incluyen medias pretest y posttest, contrastes *t* pareados, tamaños del efecto y coeficientes de fiabilidad. Los detalles completos de las pruebas de robustez se encuentran en el apéndice 3.

3.1. Resultados cuantitativos

En este subapartado se presentan los resultados derivados del análisis estadístico del pretest y posttest. En primer lugar, se examinan los cambios observados en cada uno de los ítems del cuestionario (P1-P8), con el fin de identificar patrones específicos de variación. A continuación, se analiza el índice compuesto de la competencia intercultural (P2-P8), que permite estimar el efecto global de la intervención más allá de los resultados individuales por ítem.

3.1.1. Análisis pretest-posttest por ítem (P1-P8)

Ítem	Media _{pre}	Media _{post}	Δ Media	t(59)	Valor <i>p</i>	<i>d</i> de Cohen	IC 95 % Δ Media	Shapiro-Wilk (<i>W</i> , <i>p</i>)
P1	2,75	4,97	2,22	15,32	< 0,001	1,42	[1,90; 2,54]	0,98 (<i>p</i> = 0,23)
P2	5,05	5,73	0,68	7,90	< 0,001	0,72	[0,50; 0,86]	0,97 (<i>p</i> = 0,15)
P3	4,45	5,62	1,17	11,78	< 0,001	1,12	[0,96; 1,38]	0,96 (<i>p</i> = 0,10)
P4	5,50	5,85	0,35	5,21	< 0,001	0,48	[0,15; 0,55]	0,99 (<i>p</i> = 0,40)
P5	5,10	5,68	0,58	8,74	< 0,001	0,83	[0,42; 0,74]	0,98 (<i>p</i> = 0,20)
P6	5,30	5,60	0,30	4,78	0,0002	0,41	[0,10; 0,50]	0,95 (<i>p</i> = 0,07)
P7	5,00	5,74	0,74	9,11	< 0,001	0,91	[0,58; 0,90]	0,97 (<i>p</i> = 0,14)
P8	4,15	5,30	1,15	10,92	< 0,001	1,05	[0,92; 1,38]	0,99 (<i>p</i> = 0,46)

Tabla 1. Estadísticos descriptivos y pruebas *t* pareadas ($n = 60$)

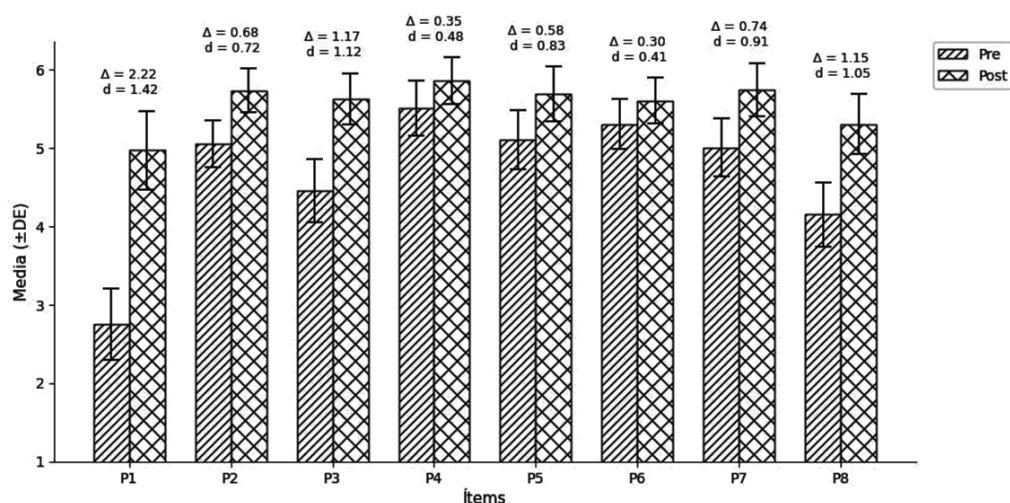


Figura 1. Gráfico de barras agrupadas: medias pretest-posttest ($\pm$ DE) por ítem

Todos los ítems muestran un aumento medio tras la sesión. Entre las barras del pretest y posttest se indican la Δ media y el d de Cohen, que varía entre 0,41 (P6) y 1,42 (P1). Estos valores, según los umbrales de Cohen (1988), presentan desde efectos pequeños a moderados hasta efectos grandes.

Los incrementos medios varían entre $\Delta \approx +0,30$ (P6) y $\Delta \approx +2,22$ (P1). Los contrastes pareados de Student muestran diferencias significativas en la mayoría de los ítems ($p < 0,001$), con la única excepción de P6 ($p = 0,0002$). La prueba de Shapiro-Wilk no rechazó la normalidad de las diferencias pretest y posttest ($p > 0,05$ para todos los ítems), por lo que la aplicación del t pareado está justificada en este conjunto de datos. La tabla 1 incluye medias del pretest y posttest, estadísticos t , valores p (ajustados y sin ajustar), tamaños del efecto (d con IC al 95 %) y medidas de fiabilidad. Los análisis complementarios (pruebas no paramétricas, remuestreo *bootstrap* y modelos mixtos) replicaron el patrón general.

3.1.2. Índice compuesto de la competencia intercultural (P2-P8)

Para valorar el efecto puntual de la sesión se construyó un índice promediando P2-P8 ($n = 60$). La tabla 2 presenta las medias del pretest y posttest, la consistencia interna (α , ω) y las métricas inferenciales principales.

El índice compuesto P2-P8 pasó de $M_{\text{pre}} = 5,08$ ($SD = 0,84$) a $M_{\text{post}} = 5,29$ ($SD = 0,42$), con una diferencia media $\Delta M = +0,21$ ($SD = 0,84$). El contraste pareado arroja $t(59) = 1,95$, $p = 0,056$; el tamaño del efecto emparejado es $d = 0,25$ (IC 95 % $d = [0,023; 0,446]$). La consistencia interna del índice aparece reflejada en $\alpha = 0,901$ y $\omega = 0,881$. La tabla 2 recoge estos y otros indicadores, cuya interpretación se desarrolla en el apartado 4, «Interpretación de los resultados».

Estadístico	Valor
Número de participantes (n)	60
Alfa de Cronbach (α) ³	0,901
Coeficiente omega (ω) ³	0,881
Media Preactividad (M_{pre})	5,08 (SD = 0,84)
Media Postactividad (M_{post})	5,29 (SD = 0,42)
Diferencia media ($\Delta M = \text{Post} - \text{Pre}$)	+0,21 (SD = 0,84)
t pareado ($t(59)$)	1,95
Valor p	0,056
Tamaño del efecto emparejado (d de Cohen)	0,25
Intervalo de confianza 95 % para d	[0,023; 0,446]
% de alumnado con mejora > 0 puntos	55,0 %
% de alumnado con mejora > 0,5 puntos	25,0 %
% de alumnado con mejora > 1,0 punto	10,0 %

Tabla 2. Resultados del índice compuesto de las preguntas P2-P8 (n = 60)

3.2. Relación con el modelo INCA: análisis por dimensiones

Para interpretar pedagógicamente los cambios observados, vinculamos los ítems del cuestionario con una versión operativa del modelo INCA centrada en tres dimensiones clave: conciencia cultural, tolerancia a la ambigüedad y flexibilidad conductual. Optamos por esta reducción porque agrupar los ítems en tres ejes facilita su interpretación, mantiene la distinción teórica entre lo cognitivo, lo actitudinal y lo procedimental y mejora la estabilidad estadística al reunir 2 o 3 ítems por dimensión.

3.2.1. Resultados por dimensión

La tabla 3 y la figura 2 recogen los resultados por dimensión operativa del modelo INCA y el índice global (P2-P8). A continuación, se destacan, de forma estrictamente descriptiva, los datos principales que aparecen en la tabla:

- Flexibilidad conductual (P3, P7): $M_{pre} = 4,90$; $M_{post} = 5,13$; $\Delta = +0,233$; $t(59) = 2,880$; $p = 0,006$; $p \text{ (BH)} = 0,024$; $d \approx 0,37$.
- Conciencia cultural (P4, P6): $M_{pre} = 5,142$; $M_{post} = 5,408$; $\Delta = +0,267$; $t(59) = 1,805$; $p = 0,076$; $d \approx 0,23$.

- Tolerancia a la ambigüedad (P2, P5, P8): $M_{pre} = 5,161$; $M_{post} = 5,322$; $\Delta = +0,161$; $t(59) = 1,298$; $p = 0,199$; $d \approx 0,17$.
- Índice global (P2-P8): $M_{pre} = 5,081$; $M_{post} = 5,293$; $\Delta = +0,212$; $t(59) = 1,946$; $p = 0,056$; $d = 0,25$.

Dimensión	M_{pre}	M_{post}	Δ	$t(59)$	p	$p(BH)$	$p(-Bonf)$	d_{Cohen}	α_{pre}	α_{post}	$r_{pre-post}$
Conciencia cultural (P4, P6)	5,142	5,408	+0,267	1,805	0,076	0,1013	0,304	0,233	0,888	0,551	0,798 / 0,386
Tolerancia a la ambigüedad (P2, P5, P8)	5,161	5,322	+0,161	1,298	0,199	0,1990	0,796	0,168	0,833	0,566	0,48–0,72 / 0,354
Flexibilidad conductual (P3, P7)	4,900	5,133	+0,233	2,880	0,006	0,0240	0,024	0,372	0,252	0,377	0,149 / 0,233
Índice global (P2-P8)	5,081	5,293	+0,212	1,946	0,056	0,1013	0,224	0,251	0,901	0,832	0,342 / 0,215

Tabla 3. Estadísticos inferenciales por dimensión operativa del modelo INCA (n = 60)

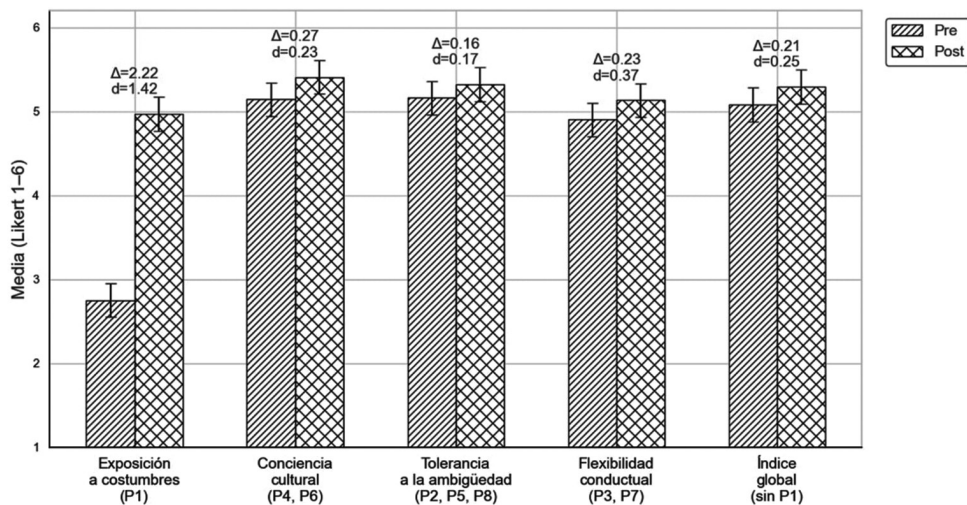


Figura 2. Comparación del pretest y posttest por dimensión del modelo INCA

La tabla 3 incluye además los valores p ajustados mediante los métodos de Benjamini y Hochberg (1995) y Bonferroni (1936), los coeficientes de consistencia interna (α , ω) y las correlaciones entre ítems. A continuación, se aborda la interpretación de estos resultados.

4. Interpretación de los resultados

Dividida en tres subapartados, se ofrece en las siguientes líneas una valoración de los resultados de este estudio. En primer lugar, se desglosa una explicación sobre los datos obtenidos al analizar cada uno de los ítems del cuestionario, así como también una apreciación del índice compuesto. En el segundo subapartado, se da cuenta de la significación, fiabilidad y robustez de los resultados. En tercer lugar, se mencionan dos aspectos a tener en cuenta a la hora de interpretar los valores recopilados: el efecto techo y la estructura de la sesión.

4.1. *Análisis ítem vs. índice compuesto*

En este diseño pre- y posttest de grupo único, tras la sesión de intervención de 100 minutos, se observa un patrón consistente de aumento en las medias de todos los ítems. Este resultado apunta a un posible efecto positivo de la intervención, aunque no permite aislar completamente su impacto, ya que no puede descartarse la influencia de las nueve semanas previas de docencia ni de otros factores contextuales. A nivel de ítem, el mayor salto se observó en P1 (conocimiento declarado), que pasó de valores bajos a una mejora notable ($\Delta \approx +2,2$; d de Cohen emparejado $\approx 1,4$). Varios ítems procedimentales y actitudinales como P3, P5, P7 y P8 presentaron también aumentos relevantes a nivel individual (véase la tabla 1).

Si agrupamos los ítems en el índice compuesto P2-P8 —diseñado para estimar el efecto inmediato de la sesión excluyendo P1— el cambio medio fue modesto pero positivo ($\Delta \approx +0,21$; d de Cohen emparejado $\approx 0,25$). Además, más de la mitad del grupo mostró alguna mejora en este índice, y una proporción apreciable mejoró medio punto o más (véase la tabla 2). Desde esta perspectiva, la magnitud de los cambios observados en las puntuaciones pretest y posttest muestra un patrón diferencial: mientras algunos ítems individuales presentan incrementos amplios, el índice compuesto refleja un cambio global más reducido, coherente con la brevedad de la intervención y con unos niveles iniciales elevados.

4.2. *Significación, fiabilidad y robustez*

La significación estadística fue elevada en la mayoría de los ítems ($p < 0,001$) y marginal en el índice global ($p = 0,056$), lo que sugiere un contraste entre mejoras puntuales y tendencias agregadas. En cuanto a la fiabilidad, la consistencia interna del índice compuesto P2-P8 fue alta ($\alpha \approx 0,90$; $\omega \approx 0,88$), aunque algunas subescalas breves (las compuestas por dos ítems) mostraron $\alpha < 0,70$ en la línea de base, lo que indica menor estabilidad inicial y por ello se aconseja cautela al interpretar efectos pequeños en esas subescalas. Por último, las comprobaciones adicionales —pruebas no paramétricas, *bootstrap* y modelos mixtos— replicaron las conclusiones principales, lo que refuerza la estabilidad de los hallazgos.

4.3. *Interpretación pedagógica y limitaciones*

Desde la perspectiva docente, cabe subrayar dos ideas prácticas. En primer lugar, destacamos un posible efecto techo, ya que muchas puntuaciones previas eran ya altas inicialmente, especialmente aquellas vinculadas a actitudes generales y conciencia cultural (por ejemplo, P2, P4, P5 y P6), con medias cercanas al extremo superior de la escala. Este punto de partida elevado reducía necesariamente el margen para observar incrementos cuantitativos amplios en estas dimensiones, de modo que cambios modestos resultan pedagógicamente relevantes. En contraste, los ítems de carácter más procedimental mostraron mayor sensibilidad al cam-

bio tras la intervención. En segundo lugar, la estructura de la sesión —que priorizó tareas procedimentales, negociación y metarreflexión— conecta directamente con las mejoras observadas en la dimensión procedimental. En particular, los ítems P3 y P7, que indagan sobre la capacidad de adaptación y la disposición a participar en interacciones interculturales, muestran la señal más consistente de cambio, lo que confirma la eficacia de actividades que promueven adaptación y práctica comunicativa.

En consecuencia, la intervención mostró eficacia inmediata, sobre todo en la dimensión de flexibilidad conductual, mientras que los cambios en conciencia cultural y tolerancia son más discretos pero relevantes dado el elevado punto de partida.

5. Discusión

Los hallazgos muestran una doble cara en la práctica docente. Por un lado, la intervención intensiva de 100 minutos se asoció con cambios detectables en múltiples indicadores de la competencia intercultural, particularmente en conductas de adaptación y disposición para la interacción. En este estudio, estas conductas se refieren a situaciones concretas como la disposición del alumnado a ajustar su comportamiento comunicativo ante interlocutores de otras culturas, participar activamente en intercambios interculturales y afrontar posibles malentendidos de manera flexible. En este sentido, los ítems P3 y P7, que exploran la capacidad percibida para adaptarse a situaciones de interacción intercultural y la voluntad de implicarse en este tipo de intercambios, muestran la señal más consistente de cambio tras la intervención. Por otro lado, cuando las mismas preguntas se agruparon en índices compuestos, la ganancia media se redujo y el índice global mostró solo una tendencia marginal ($p \approx 0,056$). En términos inferenciales, los contrastes pretest y postest fueron ampliamente significativos a nivel de ítem (véase la tabla 1) y varios ítems alcanzaron tamaños del efecto grandes; sin embargo, el índice compuesto P2-P8 exhibió una mejora media limitada y una tendencia marginal ($t(59) = 1,95$; $p = 0,056$; $d \approx 0,25$). Por tanto, recomendamos extrapolar directamente los resultados por ítem a la escala en su conjunto, pues algunos ítems pueden mejorar notablemente sin que ello se traduzca en un cambio homogéneo y consistente a nivel global.

Al desagregar por dimensiones el modelo INCA, la flexibilidad conductual emergió como la dimensión con la señal más consistente: el contraste dimensional alcanzó significación ($t(59) = 2,88$; $p = 0,006$; $p(\text{BH}) = 0,024$) y un tamaño del efecto pequeño-moderado ($d \approx 0,37$). Este resultado indica un aumento en la disposición declarada del alumnado a adaptarse y a implicarse en interacciones interculturales, tal como se recoge en los ítems del cuestionario. Conviene subrayar que esta disposición se mide a través de autopercepciones e intenciones, y no como comportamiento observable en contextos académicos reales. No obstante, desde una perspectiva pedagógica, la investigación científica ha mostrado que la percepción de autoeficacia y la disposición subjetiva influyen en la iniciativa, la implicación y la persistencia en contextos interculturales, por lo que los cambios observados pueden interpretarse como una base inicial sobre la que desarrollar prácticas interculturales más sostenidas.

En contraste, la conciencia cultural y la tolerancia a la ambigüedad registraron incrementos menores y no llegaron a la significación clásica en el análisis dimensional (véase la tabla 3), señalando que las facetas más cognitivas y actitudinales parecen requerir exposiciones mayores o instrumentos con mayor sensibilidad para evidenciar cambios sostenidos. Cabe

señalar que algunas subescalas breves, esto es, las compuestas por dos ítems, registraron $\alpha < 0,70$ en la línea de base, lo que en principio limita su sensibilidad para detectar variaciones pequeñas. No obstante, este patrón puede reinterpretarse: dado que el alfa de Cronbach es muy sensible al número de ítems, los valores bajos son esperables en escalas tan reducidas. Más relevante aún es que la consistencia interna tendió a mejorar en la medición posterior, lo que sugiere que la intervención no solo elevó las puntuaciones medias, sino que también reforzó la coherencia conceptual entre los indicadores. Desde esta perspectiva, la actividad habría funcionado como un andamiaje que ayudó a los estudiantes a conectar de manera más integrada las dimensiones evaluadas.

Consideramos que estas diferencias entre efectos puntuales y efectos agregados no son contradictorias sino informativas. Sugieren que las actividades prácticas y de negociación incluidas en la sesión —producción auténtica, contraste guiado, trabajo en grupo y metareflexión— activan con rapidez el «saber hacer» intercultural, mientras que las dimensiones más generales o cognitivas (conciencia, actitudes) dependen tanto del punto de partida del alumnado como de la capacidad discriminativa del instrumento de medición. En nuestro caso, los cambios más consistentes se concentran en los dominios vinculados a la flexibilidad conductual, concretamente en la adaptación comunicativa y la disposición a participar en interacciones interculturales (ítems P3 y P7), que se relacionan directamente con tareas de negociación de significados y contraste de perspectivas desarrolladas durante la intervención. Por el contrario, los dominios asociados a la conciencia cultural y a la tolerancia a la ambigüedad muestran incrementos más sutiles, posiblemente condicionados por un punto de partida elevado y por la menor sensibilidad de subescalas compuestas por pocos ítems, lo que reduce el margen observable de mejora a corto plazo.

Estos hallazgos encajan con marcos que distinguen capacidades procedimentales, cognitivas y actitudinales de la competencia intercultural. La mejora más consistente se observa en la flexibilidad conductual, recogida en el ítem P3, relativo a la adaptación del comportamiento comunicativo, y P7, centrada en la disposición a participar en interacciones interculturales. Este resultado respalda la idea de Byram (1997) sobre la eficacia del contraste y la práctica para activar saberes aplicados, y coincide con los enfoques CLIL que defienden la integración de contenido y lengua para movilizar destrezas (Coyle *et al.*, 2010). A su vez, la pauta observada de cambios rápidos en resultados procedimentales frente a modificaciones más lentas en actitudes se ajusta al modelo de procesos de Dear-dorff (2006), que distingue entre resultados externos observables y procesos internos de sensibilización que requieren exposiciones más prolongadas. En esta misma línea, desde el Modelo de Desarrollo de Sensibilidad Intercultural (DMIS; Bennett y Hammer, 2017) los resultados pueden interpretarse como efectos iniciales de ajuste conductual, coherentes con la idea de que las microsesiones funcionan como impulsos que favorecen la transición hacia fases adaptativas. No obstante, la progresión hacia estadios integrados requiere intervenciones reiteradas y reflexión sostenida. Finalmente, la discrepancia entre efectos puntuales y atenuación en índices compuestos refuerza la recomendación de marcos de evaluación basados en tareas como la evaluación situacional del modelo INCA. Según este, para medir fielmente la competencia intercultural, resulta conveniente combinar pruebas de desempeño y rúbricas con escalas de autoinforme, de modo que las señales rápidas de las intervenciones puedan contrastarse y respaldarse con evidencia de uso real. Para la práctica docente, estos resultados sugieren que las sesiones intensivas diseñadas en torno a tareas

auténticas y reflexión guiada pueden ser muy eficaces incluso en cursos con limitaciones de tiempo, pues permiten desarrollar la adaptabilidad y la negociación de significados en situaciones de intercambio cultural y, al mismo tiempo, aportar material inmediato para la evaluación formativa.

Al interpretar estos resultados conviene no obviar las limitaciones metodológicas que matizan su generalización. El dilatado intervalo temporal entre el pretest y la intervención (nueve semanas), combinado con la ausencia de un grupo de control y el contexto pluricultural del alumnado, implica que los cambios observados deben interpretarse como indicios exploratorios de carácter tentativo, y no como una prueba causal concluyente atribuible exclusivamente a la intervención didáctica. En particular, en este estudio no se recogieron de forma sistemática otros factores relevantes que podrían haber influido tanto en el nivel inicial como en la evolución posterior de las puntuaciones, tales como experiencias anteriores de contacto intercultural, haber crecido en entornos familiares pluriculturales, haber residido en el extranjero o haber tenido oportunidades de aprendizaje experiencial más allá del aula. Además, en el plano de la medición, ciertas subescalas alcanzaron valores de alfa de Cronbach por debajo del umbral convencional en la línea de base, lo que limita la detección de variaciones de pequeña magnitud. Asimismo, la existencia de puntuaciones iniciales elevadas en varios ítems, esto es, el efecto techo, reduce el margen cuantificable de mejora con escalas tipo Likert. Cabe añadir que la medición posterior tuvo lugar inmediatamente al terminar la sesión, de modo que desconocemos hasta qué punto los efectos se mantienen en el tiempo. Pese a estas limitaciones, las comprobaciones de robustez realizadas (pruebas no paramétricas, *bootstrap* y modelos mixtos) replicaron las conclusiones principales por ítem, lo que aporta mayor credibilidad a los cambios observados dentro del diseño aplicado. Por ello, los hallazgos deben interpretarse como indicios consistentes que invitan a ser contrastados mediante diseños más robustos —con grupos de control y seguimiento longitudinal— capaces de esclarecer la estabilidad y el alcance real de los efectos observados.

De cara al futuro, resulta recomendable combinar réplicas experimentales con un grupo de control y seguimiento longitudinal para evaluar la sostenibilidad de los efectos (Dear-dorff, 2016; Panadero y Jonsson, 2013; Fraile *et al.*, 2017). Consideramos que mejorar la instrumentación, por ejemplo, aumentando el número de ítems por dimensión, permitiría captar matices que las escalas cortas no siempre discriminan. En el caso del cuestionario aplicado, algunas dimensiones como la conciencia cultural o la tolerancia a la ambigüedad quedan representadas por un número muy reducido de ítems, lo que limita la posibilidad de diferenciar entre otros aspectos como el reconocimiento de estereotipos, la reflexión sobre la propia posición cultural o la gestión de la incertidumbre en situaciones comunicativas concretas. Complementar estas escalas con más indicadores específicos, así como con rúbricas de desempeño o tareas observacionales, contribuiría a una evaluación todavía más fina de los cambios producidos.

También resultaría imprescindible complementar los datos del cuestionario con información cualitativa, como diarios reflexivos, análisis de interacciones o producciones del alumnado. Esta combinación de fuentes permite relacionar las respuestas del cuestionario con prácticas y discursos concretos del alumnado, lo que ofrece una visión más completa y contextualizada de la competencia intercultural. De este modo, sería posible contrastar los resultados cuantitativos con lo que el alumnado hace y reflexiona en situaciones reales de clase, y comprender mejor cómo se va construyendo este aprendizaje que difícilmente puede

captarse únicamente a través de percepciones declaradas. Además, explorar la intervención en contextos menos pluriculturales permitiría analizar cómo la diversidad del aula modula el potencial de mejora (Porto, 2014 y 2016). En términos prácticos, los hallazgos apoyan la inclusión de microsesiones procedimentales en la docencia de lenguas como estrategia viable para activar competencias aplicadas, siempre acompañadas de evaluaciones más finas y réplicas que confirmen su eficacia en distintos contextos.

Conclusiones

Este estudio muestra que una intervención intensiva de 100 minutos puede producir cambios inmediatos y pedagógicamente relevantes en la competencia intercultural de estudiantes universitarios ($n = 60$), sobre todo en dimensiones procedimentales relacionadas con la adaptación y la disposición a interactuar. A nivel de ítem, se observaron incrementos claros en los indicadores de «saber hacer», especialmente en aquellos relacionados con la adaptación comunicativa y la disposición a interactuar. Estos cambios son coherentes con las tareas de la intervención basadas en la comparación de rutinas y el análisis colaborativo de choques culturales, y se reflejan de forma particular en los ítems P3 y P7. En cambio, las agregaciones en los índices compuestos ofrecieron ganancias más modestas en comparación con los cambios observados a nivel de ítem. Este patrón, corroborado por pruebas no paramétricas, remuestreos *bootstrap* y modelos mixtos, sugiere que las microsecuencias pueden activar comportamientos concretos con rapidez, aunque la transformación de actitudes o de la conciencia cultural exige exposiciones más prolongadas o instrumentos de mayor sensibilidad.

Desde la perspectiva práctica, los resultados del presente estudio avalan incluir actividades interculturales centradas en la producción real, el contraste guiado, las simulaciones de mediación y la metarreflexión incluso en cursos universitarios de ELE de nivel inicial y con escaso margen temporal para una formación específica en este ámbito. Estas secuencias facilitan la negociación de significados, fomentan la adaptabilidad comunicativa y proporcionan datos útiles para la evaluación formativa de los estudiantes al ofrecer indicios sobre las estrategias interculturales que requieren mayor apoyo o refuerzo. Asimismo, se alinean con modelos procesuales de la competencia intercultural (Deardorff, 2006) y con los enfoques de evaluación situacional del proyecto INCA, que distinguen entre resultados observables inmediatos y el desarrollo interno de la competencia en el tiempo. Para aumentar la validez y la capacidad de detección de cambios, recomendamos complementar las escalas de autoevaluación con rúbricas de desempeño, tareas observacionales y versiones multilingües de los instrumentos, así como realizar réplicas con un grupo control y seguimientos longitudinales que evalúen la sostenibilidad y la generalización de los efectos.

En definitiva, este trabajo no solo valida la utilidad de mediaciones breves para activar dimensiones aplicadas de la competencia intercultural en clases de español de niveles iniciales, sino que también ofrece un marco metodológico crítico para su evaluación, destacando la importancia de coordinar instrumentos, dimensiones y contextos en futuros estudios. La consolidación de estas ganancias puntuales en aprendizajes profundos y transferibles sigue siendo una cuestión abierta, pero los resultados aquí presentados abren una vía con potencial prometedor para integrar la interculturalidad de forma realista y efectiva incluso en condiciones curriculares restrictivas.

Referencias bibliográficas

- Benjamini, Y. y Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289-300.
- Bennett, M. J. y Hammer, M. (2017). Developmental model of intercultural sensitivity. En Y. Y. Kim (Ed.), *The International Encyclopedia of Intercultural Communication* (pp. 1-10). Wiley-Blackwell. <https://doi.org/10.1002/9781118783665.ieicc0182>
- Bilki, Z., Satar, M. y Sak, M. (2023). Critical digital literacy in virtual exchange for ELT teacher education: An interpretivist methodology. *ReCALL*, 35(1), 58-73. <https://doi.org/10.1017/S095834402200009X>
- Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R. Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3-62.
- Borghetti, C. (2017). Is there really a need for assessing intercultural competence? Some ethical issues. *Journal of Intercultural Communication*, 17(2), 1-15.
- Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.
- Cenoz, J. y Gorter, D. (2022). *Pedagogical translanguaging*. Cambridge University Press. <https://doi.org/10.1017/9781009029384>
- Cerezo, E. A. y Bertolín, J. A. (2020). Criterios para el desarrollo de las competencias interculturales en español LE/L2 desde una perspectiva crítica. *Journal of Spanish Language Teaching*, 7(2), 137-149.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Consejo de Europa. (2004). Proyecto INCA: Evaluación de la competencia intercultural.
- Consejo de Europa. (2020). *Marco común europeo de referencia para las lenguas: Aprendizaje, enseñanza, evaluación. Volumen complementario*. Estrasburgo: Servicio de publicaciones del Consejo de Europa.
- Coyle, D., Hood, P. y Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. <https://doi.org/10.1007/BF02310555>
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge. <https://doi.org/10.4324/9780203807002>
- Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, 10(3), 241-266. <https://doi.org/10.1177/1028315306287002>
- Deardorff, D. K. (2009). Implementing intercultural competence assessment. En D. K. Deardorff (Ed.), *The SAGE handbook of intercultural competence* (pp. 477-491). Sage.
- Deardorff, D. K. (2016). Outcomes assessment in international education: Changing the paradigm. En E. Jones, R. Coelen, J. Beelen y H. d. Wit (Eds.), *Global and local internationalization* (pp. 83-89). Rotterdam: Sense Publishers. https://doi.org/10.1007/978-94-6300-301-8_12
- Deardorff, D. K. (2020). *Manual for developing intercultural competencies: Story circles*. UNESCO.
- Dooley, M. y O'dowd, R. (2018). Telecollaboration in the foreign language classroom: A review of its origins and its application to language teaching practice. *In this together: Teachers' experiences with transnational, telecollaborative language learning projects*, 6, 11-34.
- Efron, B. y Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC.
- Fantini, A. E. (1995). Introduction: Language, culture and worldview: Exploring the nexus. *International Journal of Intercultural Relations*, 19(2), 143-153. [https://doi.org/10.1016/0147-1767\(95\)00025-7](https://doi.org/10.1016/0147-1767(95)00025-7)
- Fantini, A. E. (2012). Language: An essential component of intercultural communicative competence. En J. Jackson (Ed.), *The Routledge Handbook of Language and Intercultural Communication* (pp. 263-278). Routledge.

- Field, A., Miles, J. y Field, Z. (2012). *Discovering statistics using R*. Sage.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd.
- Fraile, J., Panadero, E. y Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation*, 53, 69-76. <https://doi.org/10.1016/j.stueduc.2017.03.003>
- García-Balsas, M. (2021). Interculturalidad, dimensión crítica y lengua en uso: análisis de producciones textuales de aprendices de ELE. *Tonos digital*, 41, 1-32.
- García, O. y Li, W. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.
- González Plasencia, Y. (2019). *Comunicación intercultural en la enseñanza de lenguas extranjeras*. Peter Lang.
- Gray, J. (2010). *The construction of English: Culture, consumerism and promotion in the ELT global coursebook*. Palgrave Macmillan.
- Guilherme, M. (2002). *Critical citizens for an intercultural world: Foreign language education as cultural politics*. Multilingual Matters.
- Hauck, M. (2019). Virtual exchange for (critical) digital literacy skills development. *European Journal of Language Policy*, 11(2), 187-210.
- Iglesias-Casal, I. (2021). Deconstruir y ¿evaluar? la competencia comunicativa intercultural: un desafío pendiente en la formación del profesorado. En M. Saracho-Arnáiz y H. Otero-Doval (Eds.), *Internacionalización y enseñanza del español como LE/L2: plurilingüismo y comunicación intercultural* (pp. 106-125). Oporto: Instituto Superior de Engenharia Politécnico do Porto.
- Instituto Cervantes. (2006). *Plan curricular del Instituto Cervantes: Niveles de referencia para el español*. Biblioteca Nueva.
- Kramsch, C. (1993). *Context and culture in language teaching*. Oxford University Press.
- Kumaravadivelu, B. (2003). Problematizing cultural stereotypes in TESOL. *Tesol Quarterly*, 37(4), 709-719. <https://doi.org/10.2307/3588219>
- Landone, E. (2013). Más allá de la evaluación de la competencia lingüística en la acreditación. En J. de Santiago Guervós y D. González Plasencia (Eds.), *El español global: Actas del III Congreso Internacional de Español en Castilla y León* (pp. 112-123). Junta de Castilla y León / Fundación Siglo.
- Liddicoat, A. J. y Scarino, A. (2013). *Intercultural language teaching and learning*. Wiley-Blackwell.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
- Mont, M. y Masats, D. (2018). Tips and suggestions to implement telecollaborative projects with young learners. *In this together: Teachers' experiences with transnational, telecollaborative language learning projects*, 6, 93-122.
- Panadero, E. y Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Porto, M. (2014). Intercultural citizenship education in an EFL online project in Argentina. *Language and Intercultural Communication*, 14(2), 245-261. <https://doi.org/10.1080/14708477.2014.890625>
- Porto, M. y Byram, M. (2015a). Developing intercultural citizenship education in the language classroom and beyond. *Argentinian Journal of Applied Linguistics*, 3(2), 9-29.
- Porto, M. y Byram, M. (2015b). A curriculum for action in the community and intercultural citizenship in higher education. *Language, Culture and Curriculum*, 28(3), 226-242. <https://doi.org/10.1080/07908318.2015.1087555>
- Porto, M. (2016). Ecological and intercultural citizenship in the primary EFL classroom: An online project in Argentina. *Cambridge Journal of Education*, 46(4), 395-415. <https://doi.org/10.1080/0305764X.2015.1064094>
- Precht, E. y Davidson-Lund, A. (2007). Intercultural competence and assessment: perspectives from the INCA Project. En H. Kotthoff y H. Spencer-Oatey (Eds.), *Handbook of Intercultural Communication* (pp. 467-490). De Gruyter Mouton.
- Risager, K. (2006). *Language and culture: Global flows and local complexity*. Multilingual Matters.

- Sercu, L. (2010). Assessing intercultural competence: More questions than answers? En L. Sercu y A. Paran (Eds.), *Testing the untestable in language education* (pp. 17-40). Multilingual Matters. <https://doi.org/10.21832/9781847692672-004>
- Shapiro, S. S., y Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591-611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Student. (1908). The probable error of a mean. *Biometrika*, 6(1), 1-25. <https://doi.org/10.1093/biomet/6.1.1>
- UNESCO. (2017). *Competencias Interculturales: Marco conceptual y operativo*. Universidad Nacional de Colombia. <https://unesdoc.unesco.org/ark:/48223/pf0000251592>
- Valls Campà, L. (2023). Aproximaciones a la educación para la ciudadanía intercultural y la posibilidad de su introducción en los cursos de español de las universidades japonesas. *Hispanica*, 66, 53-81.
- Van Ek, J. (1986). *Objectives for foreign language learning*. Consejo de Europa.

Notas

¹ El estadístico d de Cohen mide la magnitud de un cambio o diferencia en unidades de desviación estándar. A diferencia del valor p , que solo indica si una diferencia es estadísticamente significativa, d señala cuán grande o relevante es dicho cambio. Cohen (1988) propuso tres umbrales de referencia:

- Efecto pequeño: $d \approx 0,20$ (cambio leve, apenas perceptible en la práctica).
- Efecto moderado: $d \approx 0,50$ (cambio apreciable y con relevancia práctica).
- Efecto grande: $d \approx 0,80$ o más (cambio claro y de alta importancia práctica).

En nuestra investigación, estos valores sirven como guía para comprobar si la intervención de 100 minutos produjo un impacto menor, intermedio o considerable.

² El valor p indica la probabilidad de que los resultados observados se deban al azar. Convencionalmente, $p < 0,05$ se interpreta como evidencia suficiente para rechazar la hipótesis nula y considerar significativa la diferencia entre condiciones.

³ El alfa de Cronbach (α) y el coeficiente omega (ω) son índices de fiabilidad interna que reflejan hasta qué punto los ítems de una escala miden de manera coherente el mismo constructo. Valores $\geq 0,70$ se consideran aceptables, aunque dependen del número de ítems que componen la escala.

Perfil de los autores

Jaime Recuero Díaz es Profesor Asociado en el Instituto de Investigación en Medios y Comunicación de la Universidad de Hokkaido. Obtuvo el Máster en Formación del Profesorado con especialidad en Español como Lengua Extranjera en la Universidad de Alcalá. Ha impartido clases de español en la Universidad de Estudios Internacionales de Kobe y en la Universidad Asia Pacífico Ritsumeikan (APU). Sus intereses académicos se centran en la metodología de enseñanza del español a estudiantes asiáticos y en la investigación sobre la competencia intercultural en el aula de ELE.

Eduardo Vila López es licenciado en Periodismo por la Universidad de Navarra, posee el Certificado de Aptitud Pedagógica por la Universidad de Valencia y es máster en Enseñanza del Español como Segunda Lengua por la UNED. Desde 2008 ha trabajado como profesor de ELE en la educación secundaria y superior tanto brasileña como nipona. Asimismo, es autor de libros de texto de ELE y ha sido, durante seis años, profesor colaborador de los Institutos Cervantes de Belo Horizonte y Tokio. Entre los años 2023 y 2025 impartió clases de ELE en la Universidad Asia Pacífico Ritsumeikan (Oita). Actualmente cursa el Máster en Estudios Hispánicos Avanzados en la Universidad de Valencia, España.

Title

Assessing the intercultural competence of Spanish as a Foreign Language students: a quantitative study of a didactic intervention in a pluricultural university context

Abstract

This article explores the impact of a 100-minute intensive session designed to enhance the intercultural competence of 60 elementary-level university students of Spanish as a Foreign Language (ELE) in Japan. The intervention combined written production tasks on cultural routines, guided contrast activities, and group metareflection, assessed through an eight-item questionnaire covering knowledge, cultural awareness, tolerance of ambiguity, and behavioral flexibility. To evaluate changes, paired-samples *t*-tests were employed, complemented by nonparametric methods, bootstrap resampling, and mixed-effects models. In addition, the items were grouped into a composite index based on an adapted version of the INCA model (Council of Europe, 2004), enabling analysis of both item-level gains and overall progress. The results reveal improvements between the pretest and posttest, especially in procedural skills, and provide evidence supporting intensive didactic micro-sequences for developing intercultural competence. Future studies with control groups, longitudinal follow-ups, and mixed-methods approaches are proposed to evaluate the sustainability of effects and reinforce the instrument's psychometric validation.

Keywords

Intercultural competence, formative assessment, quantitative analysis, CLIL, Spanish as a Foreign Language (SFL).

タイトル

ELE学習者の異文化能力の評価：多文化的大学環境における教育的介入の定量的研究

要旨

本研究では、日本の大学に在籍するスペイン語学習者60名(CEFR A1レベル)を対象に、異文化間能力の育成を目的とした100分間の集中的セッションの効果を検証した。介入では、文化的ルーティンに関する記述課題、ガイド付き対比活動、グループでのメタリフレクションを組み合わせ、知識、異文化意識、あいまいさの許容度、行動的柔軟性を測定する8項目の質問紙を用いた。変化の評価には、対応のある*t*検定を中心に、ノンパラメトリック法、ブートストラップ再標本化法、混合効果モデルを併用した。また、INCAモデル(Council of Europe, 2004)を基に作成した複合指標を用いることで、個別の成果と全体的な進展の両側面から分析を行った。結果として、プレテストとポストテストの間で特に手続き的スキルにおいて改善が見られ、短時間の集中的セッションが異文化間能力の育成に有効であることが示唆された。今後は、対照群を設定した研究や縦断的追跡調査、混合研究法を取り入れ、効果の持続性を検証するとともに、測定ツールの心理計量学的妥当性を強化する必要がある。

キーワード

異文化間能力、形成的評価、定量分析、内容言語統合型学習(CLIL)、外国語としてのスペイン語(SFL)

Apéndice 1. Cuestionario (pretest y postest)

Ítem	Enunciado (Español)	Enunciado (Inglés)	Enunciado (Japonés)	Dimensión (modelo INCA)
P1	En una escala del 1 al 6: ¿cómo calificarías tu conocimiento sobre las costumbres y tradiciones de los países hispanohablantes?	On a scale from 1 to 6: how would you rate your knowledge of the customs and traditions of Spanish-speaking countries?	1～6の尺度で：スペイン語圏の国々の習慣や伝統に関するあなたの知識をどのように評価しますか。	Conocimiento declarado
P2	¿Consideras importante aprender sobre la cultura de un país al estudiar su idioma?	Do you consider it important to learn about a country's culture when studying its language?	言語を学ぶ際に、その国の文化を学ぶことは重要だと思いますか。	Tolerancia a la ambigüedad
P3	En una escala del 1 al 6: ¿cómo calificarías tu habilidad para adaptarte a normas culturales diferentes a las tuyas?	On a scale from 1 to 6: how would you rate your ability to adapt to cultural norms different from your own?	1～6の尺度で：あなたは自分と異なる文化的規範にどの程度適応できると評価しますか。	Flexibilidad conductual
P4	En mi opinión, es importante entender las costumbres y tradiciones de otros países para comunicarse efectivamente en otro idioma.	In my opinion, it is important to understand the customs and traditions of other countries in order to communicate effectively in another language.	私の意見では、他国の習慣や伝統を理解することは、別の言語で効果的にコミュニケーションするために重要です。	Conciencia cultural
P5	En mi opinión, me siento cómodo/a interactuando con personas de culturas diferentes a la mía.	In my opinion, I feel comfortable interacting with people from cultures different from my own.	私の意見では、私は自分と異なる文化を持つ人々と交流することに快適さを感じます。	Tolerancia a la ambigüedad
P6	En mi opinión, me interesa profundamente aprender sobre las tradiciones y festividades de los países hispanohablantes.	In my opinion, I am deeply interested in learning about the traditions and festivities of Spanish-speaking countries.	私の意見では、スペイン語圏の国々の伝統や祭りについて学ぶことに深い関心があります。	Conciencia cultural
P7	En mi opinión, estoy interesado/a en participar en actividades que promuevan la interacción con estudiantes de diferentes culturas.	In my opinion, I am interested in participating in activities that promote interaction with students from different cultures.	私の意見では、異文化の学生との交流を促進する活動に参加することに関心があります。	Flexibilidad conductual
P8	En mi opinión, puedo identificar similitudes y diferencias entre mi cultura y las culturas de los países hispanohablantes sin emitir juicios de valor.	In my opinion, I can identify similarities and differences between my culture and the cultures of Spanish-speaking countries without making value judgments.	私の意見では、価値判断を下すことなく、自分の文化とスペイン語圏の国々の文化との類似点や相違点を識別できます。	Tolerancia a la ambigüedad

Escala de conocimiento (P1 y P3)

- 1 = Nulo / None / 無し
- 2 = Mínimo / Minimal / ごく少
- 3 = Básico / Basic / 初歩的
- 4 = Regular / Fair / 基本的
- 5 = Bueno / Good / 良い
- 6 = Excelente / Excellent / 詳しい

Escala de acuerdo (P2, P4, P5, P6, P7, P8)

- 1 = Totalmente en desacuerdo / Strongly disagree / まったくそう思わない
- 2 = Algo en desacuerdo / Partially disagree / そう思わない
- 3 = No sé / Not sure / 分からない
- 4 = Algo de acuerdo / Somewhat agree / ややそう思う
- 5 = De acuerdo / Agree / そう思う
- 6 = Totalmente de acuerdo / Strongly agree / 非常にそう思う

Nota: En la administración del cuestionario la pregunta P1 se formuló con una indicación temporal distinta en las dos tomas: en la medición *post* se añadió «Después de la actividad de hoy y tras 3 meses estudiando español...» / «After today's activity and after three months studying Spanish...» / «本日の活動の後、かつスペイン語学習を3か月経過した時点で». Para el análisis se codificaron como P1_{pre} y P1_{post}. El resto de ítems mantuvo la misma redacción en ambos momentos.

Apéndice 2. Glosario de conceptos estadísticos empleados

Concepto	Definición	Referencias
Media (M) y Desviación estándar (SD)	La media representa el valor promedio en la escala Likert (1–6). La desviación estándar refleja cuánto varían las respuestas respecto a la media.	Field <i>et al.</i> (2012)
Δ Media	Diferencia entre la media pretest y la media posttest de cada ítem. Representa el cambio promedio observado.	—
Prueba <i>t</i> de Student para muestras apareadas	Prueba que compara medias pre y post del mismo grupo. En este estudio, <i>t</i> (59) indica 59 grados de libertad (<i>n</i> – 1).	Student (1908); Field <i>et al.</i> (2012)
Valor <i>p</i>	Probabilidad de que el cambio observado se deba al azar. Convencionalmente, <i>p</i> < 0,05 indica significación estadística.	Fisher (1925)
<i>d</i> de Cohen (apareado)	Medida del tamaño del efecto en pruebas emparejadas. Umbrales: 0,2 pequeño, 0,5 moderado, 0,8 grande.	Cohen (1988)
Intervalo de confianza (IC 95%)	Rango de valores plausibles para la estimación (ej. el tamaño del efecto). Si el IC no incluye 0, el efecto es estadísticamente significativo.	Cumming (2012)
Técnica <i>bootstrap</i>	Técnica de remuestreo (5000 réplicas en este estudio) para estimar el IC sin asumir normalidad estricta.	Efron y Tibshirani (1993)

Concepto	Definición	Referencias
Comparaciones múltiples (BH, Bonferroni)	Métodos para ajustar valores de p cuando se realizan varias pruebas simultáneamente y reducir falsos positivos.	Benjamini y Hochberg (1995)
Alfa de Cronbach (α)	Índice de consistencia interna de una escala ($\geq 0,70$ aceptable).	Cronbach (1951)
Coefficiente omega (ω)	Medida alternativa de fiabilidad de una escala, menos sensible a la cantidad de ítems.	McDonald (1999)
Shapiro–Wilk	Prueba de normalidad aplicada a las diferencias del pretest y posttest. Un $p > 0,05$ indica que no se rechaza la normalidad.	Shapiro y Wilk (1965)

Apéndice 3. Comprobaciones de robustez

Wilcoxon, tamaño del efecto (d de Cohen) con intervalos *bootstrap*, y modelo lineal mixto (*LMM*).

Ítem	W_{Wilcoxon}	p_{Wilcoxon}	d_{Cohen}	IC 95 % <i>bootstrap</i> ($n = 5000$)	θ_{LMM}	SE_{LMM}	p_{LMM}
P1	21,0	0,000	1,41	[1,12; 1,78]	—	—	—
P2	100,0	0,554	0,11	[−0,17; 0,31]	—	—	—
P3	72,0	0,007	0,37	[0,13; 0,63]	—	—	—
P4	122,5	0,053	0,27	[0,05; 0,47]	—	—	—
P5	196,0	0,419	0,15	[−0,12; 0,35]	—	—	—
P6	228,5	0,326	0,16	[−0,09; 0,38]	—	—	—
P7	106,0	0,105	0,23	[−0,02; 0,44]	—	—	—
P8	216,0	0,347	0,15	[−0,11; 0,38]	—	—	—
LMM (global)	—	—	—	—	0,40	0,07	0,000

Notas

W es el estadístico de Wilcoxon para comparación emparejada (del pretest y posttest).

El IC *bootstrap* fue calculado con 5000 réplicas sobre el d de Cohen emparejado.

LMM es el modelo mixto ajustado sobre los datos en formato largo, donde β representa el efecto fijo de la fase (del pretest y posttest), SE su error estándar asociado y p el valor p correspondiente.

Resumen interpretativo

La evidencia más robusta se observa en los ítems P1 y P3, donde la prueba de Wilcoxon es significativa y el intervalo *bootstrap* para d de Cohen excluye el valor nulo. El modelo lineal mixto global confirma un efecto positivo y significativo de la fase post, añadiendo robustez multivariada al análisis.