

ENLA 2025 Y PISA 2025: COMPARACIÓN DE MARCOS DE EVALUACIÓN, CONSTRUCTOS, CARACTERÍSTICAS DE LOS ÍTEMS Y DEMANDA COGNITIVA

Charles Anthony Arévalo Cárdenas

Resumen extendido

Las evaluaciones educativas a gran escala constituyen mecanismos fundamentales para generar evidencia sistemática sobre los aprendizajes y orientar procesos de seguimiento, análisis y toma de decisiones en los sistemas educativos. Sin embargo, la interpretación de sus resultados requiere considerar las decisiones técnicas y conceptuales que determinan qué población es evaluada, qué constructos se pretenden representar, qué referentes curriculares sustentan la evaluación, qué características poseen las tareas e ítems y qué procesos cognitivos son movilizados por los estudiantes. En este marco, el presente estudio tuvo como propósito comparar la arquitectura de evaluación de la Evaluación Nacional de Logros de Aprendizaje (ENLA) 2025, aplicada en quinto grado de secundaria en el Perú, y el Programa para la Evaluación Internacional de Estudiantes (PISA) 2025, identificando semejanzas y diferencias en cuanto al propósito de evaluación, población objetivo, constructos, referentes curriculares, marcos de evaluación, características de las tareas e ítems, contextos, formatos de respuesta y demanda cognitiva. La investigación no tuvo como finalidad comparar los resultados de desempeño obtenidos por los estudiantes ni establecer cuál de las dos evaluaciones presentó mejores resultados, sino analizar comparativamente las decisiones que sustentaron la construcción de ambas propuestas de medición. La pertinencia de este análisis se relacionó con la coexistencia de evaluaciones nacionales e internacionales dentro de los sistemas educativos contemporáneos y con la necesidad de evitar interpretaciones simplificadas basadas únicamente en la semejanza nominal de áreas como Lectura, Matemática o Ciencias. La literatura especializada sostiene que las evaluaciones a gran escala deben analizarse considerando de manera articulada el

diseño de la población, el diseño de la evaluación, los procedimientos metodológicos de análisis y el reporte de resultados (van Rijn et al., 2024). Asimismo, los enfoques contemporáneos de validez plantean que esta no constituye una propiedad aislada del instrumento, sino que se relaciona con la solidez de las interpretaciones y usos que se realizan a partir de la evidencia obtenida (Kane, 2013; Messick, 1995). En el contexto latinoamericano, Arias Ortiz et al. (2024) han señalado la coexistencia de mecanismos nacionales, regionales e internacionales de evaluación y los desafíos vinculados con su calidad, cobertura y validez. Para el contexto peruano, Damonte et al. (2017) mostraron la utilidad de examinar las correspondencias y diferencias entre evaluaciones nacionales e internacionales y los referentes establecidos por el currículo nacional.

Metodológicamente, se desarrolló una investigación cualitativa de carácter documental, descriptivo, analítico y comparativo. El corpus estuvo constituido por documentos oficiales de la Oficina de Medición de la Calidad de los Aprendizajes (UMC) y del Ministerio de Educación del Perú correspondientes a la ENLA 2025, documentos oficiales y marcos de evaluación de PISA 2025 publicados por la Organisation for Economic Co-operation and Development (OECD), así como literatura científica especializada en evaluación educativa a gran escala, validez, comparación de sistemas de evaluación, alineación curricular y características de los ítems. La unidad de análisis estuvo constituida por las características documentadas de ambas evaluaciones y no por estudiantes individuales ni por sus puntuaciones. Para organizar la información se elaboró una matriz documental integrada por veinte dimensiones: propósito, población objetivo, unidad de referencia, dominios, competencias o constructos, marco conceptual, referente curricular, conocimientos, habilidades o procesos, contextos, tipo de tarea, estímulos, formato de respuesta, modalidad de aplicación, estructura de evaluación, cobertura del contenido, demanda cognitiva, tratamiento de la dificultad, evidencias de validez y procedimientos técnicos documentados. El procedimiento comprendió la identificación y revisión de documentos oficiales, la selección de antecedentes científicos pertinentes, la extracción y sistematización de unidades de información en la matriz y el contraste de las características de ambas evaluaciones. Posteriormente, se realizó un análisis categorial de contenido orientado a identificar convergencias, divergencias y relaciones conceptuales. El rigor metodológico se fortaleció mediante triangulación documental y priorización de fuentes primarias de la UMC y OECD para la caracterización de las evaluaciones, utilizando la literatura

científica para fundamentar la interpretación conceptual y metodológica. Se excluyeron documentos de opinión, materiales promocionales sin contenido técnico y fuentes secundarias que no permitieran verificar la información. Debido a la disponibilidad pública de los materiales, el estudio no realizó una codificación exhaustiva del universo completo de ítems operativos administrados en ENLA 2025 y PISA 2025, por lo que las conclusiones sobre las características de las tareas e ítems se limitaron a aquello que se encontraba documentado en los marcos y documentos técnicos disponibles.

Los resultados evidenciaron que ENLA 2025 y PISA 2025 compartieron características propias de las evaluaciones a gran escala, entre ellas la definición de poblaciones, dominios de evaluación, constructos, especificaciones de tareas, procedimientos estandarizados y mecanismos técnicos destinados a generar evidencia sobre conocimientos y habilidades. No obstante, se identificaron diferencias sustantivas en la población objetivo, propósito, referentes de construcción, amplitud de los dominios y configuración de las tareas. La ENLA 2025 se articuló explícitamente con los referentes curriculares nacionales y con las competencias previstas para la educación básica peruana. En quinto grado de secundaria consideró Lectura, Matemática, Ciencia y Tecnología y Desarrollo Personal, Ciudadanía y Cívica. Su organización respondió a grados específicos de la educación básica peruana y, por tanto, su población de referencia estuvo definida principalmente desde la estructura del sistema educativo nacional. PISA 2025, en cambio, estableció su población internacional principalmente mediante un criterio etario, dirigido a estudiantes de aproximadamente 15 años escolarizados, independientemente del grado específico en que se encontraran. Esta diferencia implica que, aunque pudiera existir un solapamiento de edades entre determinados estudiantes de quinto grado de secundaria y participantes de PISA, ambas poblaciones no fueron conceptualmente equivalentes. En relación con los marcos de evaluación, ENLA se construyó a partir de las competencias y referentes curriculares correspondientes a la educación básica peruana, mientras que PISA utilizó marcos internacionales destinados a evaluar la movilización y aplicación de conocimientos y habilidades en situaciones contextualizadas, sin reproducir íntegramente el currículo de un país determinado. PISA 2025 incorporó, además, Ciencias como dominio principal y el dominio innovador Learning in the Digital World, junto con Lectura y Matemática y una evaluación opcional de inglés como lengua extranjera.

Respecto de las tareas, estímulos e ítems, ambas evaluaciones utilizaron situaciones diseñadas para obtener evidencia sobre conocimientos y habilidades, pero estas se organizaron conforme a marcos conceptuales diferentes. En ENLA, las características de textos, preguntas y tareas se vincularon con las competencias y áreas establecidas en las especificaciones nacionales. En PISA, las tareas se estructuraron alrededor de situaciones contextualizadas que exigieron movilizar conocimientos y procesos definidos por los respectivos marcos de dominio. La comparación mostró, además, que el formato de respuesta constituye una dimensión relevante de la arquitectura de evaluación. Marcq et al. (2024), a partir del análisis de aproximadamente 260 000 estudiantes de 71 países participantes en PISA 2018, encontraron diferencias sistemáticas entre ítems de respuesta construida y opción múltiple en términos de dificultad. Este antecedente permitió sostener que el formato no debe considerarse una característica superficial, debido a que puede relacionarse con las condiciones bajo las cuales se produce la evidencia de aprendizaje. En cuanto a los contextos, se identificó su presencia en ambas evaluaciones, aunque con finalidades y referentes diferentes. PISA 2025 incorporó explícitamente contextos personales, locales o nacionales y globales en determinados dominios, mientras que ENLA articuló la contextualización con las competencias y situaciones previstas en sus documentos nacionales. Sin embargo, la existencia de un contexto no implicó por sí misma una mayor demanda cognitiva, puesto que la complejidad dependió de las operaciones requeridas, los conocimientos movilizados y las relaciones que el estudiante debía establecer.

Uno de los hallazgos centrales correspondió a la diferenciación entre constructo, formato de respuesta, demanda cognitiva y dificultad. El constructo se relacionó con aquello que la evaluación pretendió representar; el formato determinó la modalidad mediante la cual el estudiante produjo su respuesta; la demanda cognitiva se vinculó con el tipo de procesamiento mental requerido para resolver una tarea; y la dificultad correspondió a una característica empírica asociada con el comportamiento observado de los estudiantes frente al ítem. En consecuencia, no resultó metodológicamente apropiado inferir automáticamente la demanda cognitiva de una pregunta a partir de verbos como identificar, analizar, interpretar o evaluar, ni asumir que una tarea con determinado proceso cognitivo debía ser necesariamente fácil o difícil. El marco de PISA 2025 distingue expresamente la demanda cognitiva de la dificultad empírica, esta última relacionada con la proporción de estudiantes que logra resolver correctamente un

ítem. Por ello, una tarea puede presentar determinadas exigencias cognitivas y, simultáneamente, mostrar diferentes niveles de dificultad empírica dependiendo de las características de los estudiantes y de la interacción entre conocimientos, procesos y condiciones de la tarea.

La discusión permitió establecer que la comparación entre ENLA 2025 y PISA 2025 resulta pertinente cuando se aborda desde la arquitectura de evaluación y no desde una comparación directa de resultados. La diferencia entre sus poblaciones constituye una condición fundamental para determinar qué inferencias pueden realizarse. Del mismo modo, la diferencia entre un referente curricular nacional y un marco internacional condiciona la definición de los constructos y la selección de contenidos, procesos y contextos. Estos hallazgos son consistentes con van Rijn et al. (2024), quienes plantearon la necesidad de considerar las características del diseño de evaluación antes de establecer comparaciones entre sistemas. También guardan relación con Damonte et al. (2017), al mostrar que una evaluación nacional y una internacional pueden presentar distintos grados de correspondencia con el currículo peruano. Desde la perspectiva de la validez, las diferencias identificadas no permiten establecer que una evaluación sea intrínsecamente más válida que la otra, dado que la validez depende de las evidencias disponibles para sustentar las interpretaciones y usos propuestos de cada sistema de medición (Kane, 2013). En este sentido, ENLA proporciona evidencia vinculada con el sistema educativo peruano y sus referentes curriculares nacionales, mientras PISA genera evidencia bajo un marco internacional común orientado a permitir comparaciones entre diferentes sistemas educativos. La principal contribución del estudio consiste, por tanto, en mostrar que dos evaluaciones que comparten determinados dominios pueden responder a lógicas de construcción diferentes y producir evidencias que no deben considerarse automáticamente equivalentes.

En conclusión, ENLA 2025 y PISA 2025 pudieron ser comparadas de manera analítica respecto de sus marcos de evaluación, constructos, referentes curriculares, población objetivo, tareas, ítems, contextos y demanda cognitiva, pero no debieron tratarse como evaluaciones intercambiables. Ambas compartieron una orientación hacia la generación sistemática de evidencia sobre conocimientos y habilidades, pero se diferenciaron en la definición de sus poblaciones, propósitos, referentes curriculares, amplitud de dominios y arquitectura de las tareas. ENLA se vinculó explícitamente con las competencias y referentes curriculares de la educación básica peruana, mientras PISA se estructuró

desde marcos internacionales centrados en la aplicación de conocimientos y habilidades en situaciones contextualizadas. Asimismo, la demanda cognitiva no pudo determinarse únicamente mediante el formato de los ítems o los verbos empleados en las consignas, debido a que dependió de la interacción entre procesos, conocimientos y características específicas de las tareas. El estudio aporta una base conceptual y metodológica para futuras investigaciones orientadas a comparar empíricamente ítems liberados de ambas evaluaciones mediante procedimientos de codificación independiente y análisis de concordancia entre evaluadores. De esta manera, se fortalece una lectura técnicamente fundamentada de las evaluaciones educativas y se evita reducir la comparación entre ENLA y PISA a sus resultados numéricos, reconociendo que toda evidencia de evaluación adquiere significado dentro del diseño, población, constructo y propósito para los cuales fue producida.

Palabras clave

ENLA 2025; PISA 2025; evaluación educativa; evaluaciones a gran escala

Referencias:

Arias Ortiz, E., Dueñas, X., Giambruno, C., & López, Á. (2024). The state of education in Latin America and the Caribbean: Learning assessments. Inter-American Development Bank. https://doi.org/10.18235/0013171

Damonte, G., Cabrera, A., & Miranda, F. (2017). Análisis de confluencia de enfoques y expectativas de aprendizaje entre las evaluaciones de rendimiento (internacionales y nacionales) y el Currículo Nacional, en relación al tema ambiental. GRADE. https://hdl.handle.net/20.500.12820/532

Kane, M. T. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1–73. https://doi.org/10.1111/jedm.12000

Marcq, S., Donayre, L., & Braeken, J. (2024). The role of item format in the PISA 2018 mathematics literacy assessment: A cross-country study. Studies in Educational Evaluation, 83, 101401. https://doi.org/10.1016/j.stueduc.2024.101401

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning.

American Psychologist, 50(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Organisation for Economic Co-operation and Development. (2026). PISA 2025 assessment and analytical framework. OECD Publishing. https://doi.org/10.1787/86c36975-en

Oficina de Medición de la Calidad de los Aprendizajes. (2026). Reporte técnico de la Evaluación Nacional de Logros de Aprendizaje de Estudiantes (ENLA) 2025. Ministerio de Educación del Perú.

van Rijn, P., Por, H.-H., McCaffrey, D. F., Bhaduri, I., & Bertling, J. (2024). A framework for comparing large-scale survey assessments: Contrasting India's NAS, United States' NAEP, and OECD's PISA. Frontiers in Education, 9, 1422030. https://doi.org/10.3389/feduc.2024.1422030