



<https://doi.org/10.18222/dae.v35.10997>

ANÁLISIS COMPARATIVO DE MATRICES DE EVALUACIÓN: ENEM, PAEBES, PLANEA Y EXEIMS-BC

 MATEUS GOBBI DOS SANTOS^I

 RONILDO STIEG^{II}

 BERENICE MORALES GONZÁLEZ^{III}

 WAGNER DOS SANTOS^{IV}

^I Universidade Federal do Espírito Santo (UFES), Vitória-ES, Brasil; gobbimateus2001@gmail.com

^{II} Universidade Federal do Espírito Santo (UFES), Vitória-ES, Brasil; ronildo.stieg@yahoo.com.br

^{III} Benemérita Escuela Normal Veracruzana "Enrique C. Rébsamen" (BENV), Xalapa-VE, México; berenice_mg@yahoo.com

^{IV} Universidade Federal do Espírito Santo (UFES), Vitória-ES, Brasil; wagnercefd@gmail.com

RESUMEN

Este artículo tiene como objetivo analizar y comparar las matrices de referencia de dos políticas de evaluación nacionales (Enem/Brasil y Planea/México) y dos estatales (Paebes/Espírito Santo y Exeims-BC/Baja California). El carácter cualitativo y documental se basa en el método comparativo, utilizando el *software* Iramuteq, y las fuentes son las matrices de referencia de los cuatro exámenes. Los resultados apuntan a la similitud en los componentes curriculares evaluados, con énfasis en matemáticas y lengua materna, además del predominio de competencias relacionadas con el ámbito cognitivo. En conclusión, se cuestiona la relevancia de dos evaluaciones similares en componentes curriculares y competencias evaluadas para la educación secundaria.

PALABRAS CLAVE POLÍTICAS DE EVALUACIÓN • ESCUELA SECUNDARIA • AMÉRICA LATINA • MATRICES DE REFERENCIA.

COMO CITAR:

Santos, M. G. dos, Stieg, R., González, B. M., & Santos, W. dos. (2024). Análisis comparativo de matrices de evaluación: Enem, Paebes, Planea y Exeims-BC. *Estudos em Avaliação Educacional*, 35, Artículo e10997. <https://doi.org/10.18222/dae.v35.10997>

ANÁLISE COMPARATIVA DE MATRIZES DE AVALIAÇÃO: ENEM, PAEBES, PLANEA E EXEIMS-BC

RESUMO

Este artigo tem como objetivo analisar e comparar as matrizes de referência de duas políticas de avaliação nacionais (Enem/Brasil e Planea/México) e duas estaduais (Paebes/Espírito Santo e Exeims-BC/Baixa Califórnia). O caráter qualitativo e documental baseia-se no método comparativo, utilizando o *software* Iramuteq, e as fontes são as matrizes de referência dos quatro exames. Os resultados apontam a similaridade dos componentes curriculares avaliados, com ênfase nas matemáticas e língua materna, além do predomínio de competências relacionadas com o âmbito cognitivo. Em conclusão, questiona-se a relevância de haver duas avaliações similares em componentes curriculares e as competências avaliadas para o ensino secundário.

PALAVRAS-CHAVE POLÍTICAS DE AVALIAÇÃO • ENSINO MÉDIO • AMÉRICA LATINA • MATRIZES DE REFERÊNCIA.

COMPARATIVE ANALYSIS OF EVALUATION MATRIXES: ENEM, PAEBES, PLANEA AND EXEIMS-BC

ABSTRACT

The purpose of this article is to analyze and compare the reference matrices of two national evaluation policies (Enem/Brazil and Planea/Mexico) and two state policies (Paebes/Espírito Santo and Exeims-BC/Baja California). This study is qualitative and documentary in nature. It is based on the comparative method, using the Iramuteq software, and as sources, the reference matrices of the four exams. The results point to the similarity in the curricular components evaluated, with emphasis on mathematics and mother tongue, in addition to the predominance of competencies related to the cognitive area. In conclusion, the relevance of two similar evaluations in terms of curricular components and competencies assessed for secondary education is questioned.

KEYWORDS ASSESSMENT POLICIES • HIGH SCHOOL • LATIN AMERICA • REFERENCE MATRIXES

Recebido el: 7 MARZO 2024

Aprobado para publicación el: 6 AGOSTO 2024



Este es un artículo de acceso abierto que se distribuye en los términos de la licencia Creative Commons, tipo BY-NC.

INTRODUCCIÓN

Las evaluaciones educativas a gran escala, también conocidas como exámenes o pruebas estandarizadas, son evaluaciones sistémicas diseñadas para medir el logro de objetivos educativos trazados para estudiantes en edad escolar, en uno o más componentes curriculares, de un determinado sistema educativo (municipal, estatal, nacional o internacional) cuyas características, aplicación y corrección tienden a ser uniformes (Mattos Santos et al., 2024). Sus resultados promueven escalas comparativas entre sistemas educativos, indicando posibles factores contextuales del sistema educativo (escuela, estrategias pedagógicas y acción docente) en el nivel de desempeño de los estudiantes en estas evaluaciones (Mons, 2009).

Según Carrasco y Gunter (2019), las evaluaciones educativas a gran escala tienen un enfoque gerencialista impulsado por las reformas educativas a nivel global, justificándose en el discurso de que los datos que generan alientan estrategias para mejorar la calidad de la educación. Sin embargo, los autores afirman que este movimiento tuvo un impacto considerable en la organización de los sistemas educativos, especialmente en países con sistemas escolares organizados en cuasimercados, como Chile (Carrasco & Gunter, 2019).

Se entiende que el estudio de las evaluaciones educativas a gran escala requiere considerarlas como acciones basadas en políticas educativas, que tienden a servir como mecanismos regulatorios impulsados por organismos gubernamentales y agentes externos contratados por ciertos gobiernos u organismos multilaterales (Monarca, 2015; Dubet, 2016; Parra & Matus, 2016; Ozga, 2020). Dadas estas características, Bauer et al. (2015) llaman la atención sobre el hecho de que las evaluaciones educativas a gran escala han asumido cada vez más roles protagónicos en las políticas educativas mediante la creación y expansión de programas de evaluación específicos en diferentes contextos (Uczak, 2014; Casaril, 2016; Álvarez-López & Matarranz García, 2020).

Estudios como el de Mattos Santos et al. (2023) indican que los países latinoamericanos cuentan con un sistema nacional de evaluación de la educación básica. La justificación que presentan estos sistemas consiste en afirmar que este tipo de evaluación es fundamental para orientar las políticas educativas, pudiendo abarcar los ámbitos estatal, nacional e internacional (Bauer et al., 2015). En este escenario, hay evidencia del surgimiento de políticas de evaluación también a nivel local, relacionadas con el proceso de rendición de cuentas en educación, que Baker (2009) llama una forma de responsabilizar a estudiantes, docentes y equipos pedagógicos por sus resultados en las evaluaciones a gran escala mientras se intenta alcanzar objetivos educativos específicos. Para Linn (2008), el objetivo de un sistema de rendición de cuentas es distinguir a las escuelas en términos de eficacia

docente, de modo que se puedan tomar medidas para mejorar el desempeño de los estudiantes en escuelas consideradas ineficaces.

Considerando la rendición de cuentas como factor determinante en el desarrollo de cada vez más programas de evaluación (exámenes/pruebas) de gran escala dentro de los sistemas educativos, especialmente en las etapas de enseñanza obligatoria, suelen estar referenciados por criterios basados en estándares curriculares estatales, junto con estándares de exámenes nacionales e internacionales de referencia (Álvarez-López & Matarranz García, 2020). En este sentido, los Estados, en sus estrategias políticas, con el objetivo de lograr que sus estudiantes obtengan buenos desempeños en los resultados de las evaluaciones educativas de gran escala, han fomentado la práctica de la preparación constante para exámenes estandarizados.

Miras y Solé (1996) señalan que, a la vista de los objetivos educativos estipulados y la necesidad de conocer su alcance, se justifica la necesidad de medir y analizar la situación educativa a través de procesos de evaluación, que pueden considerarse como una actividad a través de la cual, en función de determinados criterios, se obtiene información pertinente sobre un fenómeno, situación, objeto o persona, se emite un juicio sobre el objeto en cuestión y se adoptan una serie de decisiones relativas a este (Miras & Solé, 1996).

Para garantizar importancia, validez, confiabilidad y precisión, las evaluaciones educativas a gran escala se basan en metodologías psicométricas de análisis de datos. Por su alcance, estas evaluaciones suelen estar reguladas por leyes u ordenanzas, dependiendo del modelo político-organizativo y el contexto en el que se producen. Así, es posible generar a partir de este proceso documentos capaces de funcionar como guías normativas y/o matrices de referencia que pretenden regular y orientar el tratamiento de las evaluaciones educativas de gran escala.

En Brasil, la Lei de Diretrizes e Bases da Educação Nacional (LDB) (Lei n. 9.394, 1996) señala el papel de la evaluación a gran escala, requiriendo su universalización en todo el territorio nacional. La propuesta consiste en una evaluación definida como una prioridad asociada al desarrollo de sistemas de información y evaluación en todos los niveles y modalidades de educación, con el fin de mejorar los procesos de recolección y difusión de datos y mejorar la gestión y mejorar la enseñanza (Werle, 2011).

Así lo ha reforzado el Plano Nacional de Educação [Plan Nacional de Educación] (PNE), aprobado en 2014 y vigente hasta 2024, sobre la necesidad de establecer sistemas de evaluación a gran escala. Según Ferreira et al. (2017), en el PNE se observa el protagonismo otorgado a evaluaciones de esta naturaleza, junto con el Índice de Desenvolvimento da Educação Básica [Índice de Desarrollo de la Educación Básica] (Ideb), como principales indicadores de la calidad de la educación

básica, sirviendo al mismo tiempo como instrumentos de seguimiento de la calidad de la educación básica y los lineamientos de las políticas públicas.

Acciones similares se identifican en México con base en la Ley General de Educación, del 13 de julio de 1993 (LGE), que establece una evaluación del sistema educativo, bajo la responsabilidad de la Secretaría de Educación Pública, sin generar daños a las evaluaciones desarrolladas por las secretarías de cada entidad federativa. En este caso, las evaluaciones tanto nacionales como estatales, presentes en el territorio mexicano, deben ser sistemáticas y permanentes, tomando sus resultados como base para la adopción de medidas en el nivel de las autoridades competentes (González & Bosco, 2006). La citada ley también establece que las instituciones educativas deberán colaborar con estas evaluaciones, brindando la información necesaria en tiempo oportuno y apoyando los procesos de evaluación. Según González y Bosco (2006), la LGE recomienda a las autoridades educativas dar a conocer los resultados de las evaluaciones a docentes, estudiantes, padres de familia y a la sociedad en general, con el objetivo de comprender el desarrollo y avances de la educación en cada unidad escolar.

En un escenario en el que se promulgan leyes nacionales para establecer evaluaciones educativas de gran escala, con el discurso de seguimiento y seguimiento de las metas proyectadas para los sistemas educativos, ejemplos como el Exame Nacional do Ensino Médio (Enem) y el Programa de Avaliação da Educação Básica do Espírito Santo [Programa de Evaluación de Educación Básica en Espírito Santo] (Paebes) en Brasil, así como el Plan Nacional de Evaluación de los Aprendizajes (Planea) y el Examen Estatal de Ingreso a Media Superior (Exeims-BC) en Baja California en México. Llama la atención que se elaboraron guías (Planea), tablas (Exeims-BC) y matrices de referencia (Enem y Paebes) para orientar estas evaluaciones a nivel nacional y estatal en ambos países. Estos documentos sirven como referencia curricular, indicando lo que se evaluará en cada componente curricular y grado/año escolar, así como las competencias y habilidades que se esperan que sean alcanzados por los estudiantes. Es importante mencionar que, en este estudio, se optó por el concepto de matrices de referencia como forma de estandarizar conceptos de los cuatro documentos.

Por lo tanto, surgen interrogantes sobre estas evaluaciones, tales como: ¿Cuáles son los componentes curriculares y las competencias señaladas en las matrices de referencia de Enem, Paebes, Planea y Exeims-BC? ¿Existen diferencias en los componentes curriculares y en las competencias evaluadas en los exámenes nacionales (Enem y Planea) y estatales (Paebes y Exeims-BC) en Brasil y México, respectivamente, dirigidos a estudiantes de secundaria? ¿Qué aproximaciones hay en estas políticas de evaluación educativa en diferentes contextos, considerando las peculiaridades de Enem, Paebes, Planea y Exeims-BC?

A partir de estas preguntas, el objetivo de este artículo es analizar y comparar las matrices de referencia de Enem, Paebes, Planea y Exeims-BC, identificando y discutiendo los componentes curriculares y las competencias perfiladas en estos documentos. Además, se busca comprender si existen similitudes y/o diferencias entre estas políticas de evaluación. Cabe destacar que la evaluación educativa de gran escala Enem cubre todo el territorio brasileño, mientras que Paebes se aplica en el estado de Espírito Santo. Planea cubre todo el territorio mexicano y Exeims-BC es adoptado por el estado de Baja California. Todos estos exámenes están destinados a estudiantes de secundaria.

MARCO TEÓRICO

Las reformas educativas impulsadas en los años 1990 en los países latinoamericanos coincidieron con una fase de consolidación y expansión de las evaluaciones educativas internacionales a gran escala, impulsadas por la Asociación Internacional para la Evaluación del Desempeño Educativo (AIE) y la Organización para la Cooperación y el Desarrollo Económico (OCDE), con el objetivo de desarrollar sistemas de evaluación y la participación de algunos países en los principales estudios internacionales (Elías, 2017). En el mismo contexto, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco) también jugó su influyente papel al impulsar la creación del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE).

Según Elías (2017), el énfasis puesto en el discurso evaluativo como aspecto central de las políticas educativas se ha basado principalmente en pasar del acceso a la educación a medir los aprendizajes que logran los estudiantes con base en estándares internacionales. Esto implica valorar los estudios de evaluación estandarizados, especialmente las comparaciones internacionales, como el Programa para la Evaluación Internacional de Estudiantes (PISA) impulsado por la OCDE. PISA, a su vez, se convierte en una política educativa global que modifica el papel del Estado en relación con sus atribuciones como definidor y agente responsable de la implementación de las políticas educativas, teniendo influencias de otros organismos no gubernamentales, incluidas empresas multinacionales vinculadas a la producción y venta de servicios educativos (Verger et al., 2012; Sellar & Lingard, 2013; Volante, 2015; Baird et al., 2016). Además, Verger et al. (2017) señalan que la globalización afecta drásticamente las políticas educativas de los países latinoamericanos, ya que introducen mecanismos de mercado en el sector educativo a través de la privatización y la importación de tecnologías de enseñanza, con el objetivo de obtener mejores resultados en evaluaciones a gran escala.

Ante este escenario, se identifica la existencia de tendencias en las políticas educativas globales, como una influencia creciente de las evaluaciones educativas a gran escala en las políticas educativas de los diferentes países, en lo que respecta al diseño de programas de evaluación y la propuesta de proyectos educativos para mejorar la calidad de la educación (Elías, 2017). En consecuencia, esto ha generado agendas educativas que han negado las diferencias culturales, apuntando a la convergencia de indicadores únicos y al desarrollo de *rankings* entre países y sistemas educativos, como un discurso de valoración y búsqueda del progreso en educación.

Considerando que las evaluaciones educativas internacionales a gran escala afectan las políticas educativas, este estudio busca, a través del análisis documental crítico y el método comparativo, comprender las posibles influencias de las evaluaciones educativas nacionales de gran escala (Enem, en Brasil, y Planea, en México) en las evaluaciones estatales (Paebes, en Espírito Santo, y Exeims-BC, en Baja California). Según Bloch (2001), el análisis documental crítico consiste en no realizar una lectura superficial de las fuentes, pues existe la posibilidad de obtener respuestas insuficientes cuando no se cuestionan. Esto se debe a que los documentos pueden haber sido producidos con diferentes intenciones, en función de necesidades o demandas específicas y en contextos variados, ya sean políticos, regionales o culturales, y en diferentes períodos cronológicos.

El análisis documental crítico asociado al método comparativo (Bloch, 1998) refuerza que los elementos de una fuente deben ser comprendidos en sus particularidades y desde su contexto, para resaltar las características comparables y las incomparables. De esta manera, los análisis comparativos tuvieron en cuenta las especificidades de cada contexto (evaluaciones a gran escala) y sus posibles aproximaciones (Bloch, 1998), a partir de una misma política educativa.

En el ámbito de la evaluación educativa, es importante señalar que las evaluaciones educativas a gran escala se basan principalmente en metodologías psicométricas de análisis de datos, área de estudio relacionada con técnicas de medición psicológica, incluyendo la medición de conocimientos, habilidades, actitudes, rasgos de personalidad y medición educativa (Pasquali, 2017; Bazán, 2018). Cabe destacar que las metodologías psicométricas de análisis de datos incluyen la teoría clásica de los test (TCT) y la teoría de respuesta al ítem (TRI). La TCT se puede utilizar en diferentes contextos sin implementar supuestos rigurosos; en este caso, se supone que la puntuación de una prueba es, por naturaleza, defectuosa y contiene algún error. En consecuencia, el resto debe tener una base real o verdadera (Fletcher, 2010). La TRI, ampliamente difundida en las últimas décadas en el escenario de las evaluaciones educativas a gran escala, consiste en la recopilación de modelos estadísticos utilizados para hacer predicciones, estimaciones o inferencias sobre

las habilidades (o competencias) medidas en una prueba. Además, permite analizar datos desde una característica más cualitativa (Rocha, 2018).

La disparidad entre las dos metodologías psicométricas de análisis de datos pone de relieve la relación entre la dificultad de los ítems y las puntuaciones obtenidas. En TCT, la puntuación del individuo está sujeta a fluctuaciones, dependiendo de si la prueba aplicada es más fácil o más difícil, lo que resulta en variaciones en errores y puntuaciones. En cambio, en la TRI, independientemente de los ítems o conjunto de ítems utilizados, siempre que estén destinados a medir el mismo rasgo latente, conducirán al mismo nivel de aptitud del sujeto. Esto ocurre, por supuesto, considerando los inevitables errores de medición presentes en cualquier dominio científico (Pasquali & Primi, 2003).

Otro concepto movilizado en este estudio se refiere a las competencias, definidas por Zabala (1998) como la capacidad del sujeto para movilizar conocimientos, habilidades y actitudes para resolver problemas y tomar decisiones adecuadas. Según Perrenoud (1999), las competencias son efectos adaptativos del ser humano para resolver problemas o superar una situación. Según el autor, las competencias requieren conocimientos/contenidos para ser efectivas, especialmente aquellas que se desarrollan en el contexto escolar.

A partir de este marco teórico y teniendo en cuenta su carácter exploratorio, el presente estudio expresa su importancia en el campo de la evaluación educativa, ya que surge de la necesidad de analizar las matrices de referencia de las evaluaciones educativas de gran escala en cuestión, en busca de comprender similitudes y distancias entre documentos en sus respectivos contextos (Phillips & Ochs, 2003; Steiner-Khamsi, 2014, 2016).

METODOLOGÍA

De carácter cualitativo y exploratorio (Creswell & Clark, 2013), este estudio se basa en el análisis documental crítico (Bloch, 2001) y el método comparativo (Bloch, 1998). Las fuentes estuvieron constituidas por las matrices de referencia de cuatro evaluaciones educativas a gran escala con diferentes niveles de cobertura territorial: Enem (Brasil) y Planea (México) a nivel nacional, y Paebs (Espírito Santo/Brasil) y Exeims-BC (Baja California/México) a nivel estatal.

Este estudio es el resultado de un acuerdo de colaboración en investigación entre la Universidade Federal do Espírito Santo (Brasil) y Universidad Benemérita Escuela Normal Veracruzana “Enrique C. Rébsamen” (México). Los datos de esta investigación están vinculados a dos proyectos con financiación: “Avaliação estandardizada nos estados do Espírito Santo/Brasil e Baja California/México: dilemas e tensões do Paebs e do Exeims-BC”, Edital Universal Fapes (Fundação

de Amparo à Pesquisa e Inovação do Espírito Santos) n. del proceso 11367/2021, y “Dos exames estandarizados ao direito à aprendizagem: experiências para a melhoria da qualidade do ensino médio na Rede Estadual do Espírito Santo”, Edital Universal Fapes n. 28/2022. Las fuentes fueron obtenidas a través de los sitios web de los Ministerios de Educación de Brasil y México y de las Secretarías de Educación de los estados de Espírito Santo y Baja California, siendo los principales órganos responsables de la elaboración de los documentos normativos de estos exámenes, asumidos en este estudio como matrices de referencia de Enem (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [Inep], 2010), Planea (Instituto Nacional para la Evaluación de la Educación [Inee], 2019), Paebes (Secretaria de Estado da Educação do Espírito Santo [Sedu], 2022) y Exeims-BC (Colégio de Bachilleres del Estado de Baja California [CoBachBC], 2020), respectivamente.

Para tabular y analizar los datos se utilizó como herramienta de ayuda el *software* Interface de R pour les Analyses Multidimensionnelles de Textes et Questionnaires (Iramuteq), versión 0.7 Alpha 2. Se utilizó Iramuteq para generar las nubes de palabras (figuras 1, 2, 5 y 6) y las similitudes de palabras (figuras 3, 4, 7 y 8). En el estudio se optó por el análisis textual a través de la interfaz de lexicografía básica (nubes) y coocurrencias entre palabras (similitud), una representación visual generada como resultado de las relaciones de conexión entre palabras. Es importante resaltar que, si bien se trata de fuentes de contextos distintos (Brasil y México) que tienen idiomas portugués y español, respectivamente, las figuras 1 a 4 producidas en Iramuteq tuvieron en cuenta el contenido original de los documentos. El objetivo fue preservar la integridad de los documentos brasileños tal como están disponibles, ya que no existen versiones en español.

Después de reunir todas las matrices de referencia, se produjeron cuatro archivos en formato txt (bloc de notas), uno para cada matriz. Estos archivos luego fueron enviados individualmente al Iramuteq, generando las figuras 1 a 8. Se consideraron palabras con recurrencia \geq tres en cada categoría, determinada por el propio *software*. La elección de analizar datos textuales se debió a su capacidad de agrupar, relacionar y comparar, a través de representaciones gráficas basadas en el análisis léxico (Camargo & Justo, 2013), prevalente en las matrices de referencia. Cabe señalar que el *software* no es un método y las imágenes generadas por él no representan, en sí mismas, el análisis de datos, siendo sólo un instrumento de exploración. La interpretación de estos resultados se queda a cargo del investigador, quien debe utilizar conocimientos externos al texto (Chartier & Meunier, 2011).

La interpretación de los datos se realizó a través del análisis comparativo de fuentes, considerado una práctica esencial en la investigación cualitativa, ya que evita un enfoque unilateral en el objeto investigado (Bloch, 1998). Con los resultados del *software* fue posible identificar qué es lo más recurrente en las matrices de

referencia, especialmente en lo que respecta a los componentes curriculares y las competencias evaluadas y esperadas de los estudiantes de secundaria en estas evaluaciones. Además, buscamos identificar las metodologías psicométricas de análisis de datos que subyacen a las evaluaciones educativas a gran escala en los contextos analizados.

RESULTADOS Y DISCUSIÓN

Para presentar las características de las evaluaciones educativas a gran escala seleccionadas para este estudio, se elaboró la Tabla 1. En ella se indican las siglas de las evaluaciones, el año de implementación, el año de la matriz de referencia utilizada en este estudio, el alcance (nacional y/o estatal) y las áreas de conocimiento y componentes curriculares evaluados.

Características generales de las evaluaciones educativas a gran escala

A partir de la Tabla 1, se identifican cuatro evaluaciones educativas a gran escala ubicadas en diferentes regiones geopolíticas: América del Norte (México) y América del Sur (Brasil). Además, se ubican dos estados (Espírito Santo y Baja California), que también cuentan con un examen estandarizado a nivel local. La elección de estos contextos se basó en la idea de que estos dos países tienen una organización geopolítica de estados federados, lo que permite cierta autonomía a estos estados en la delimitación e implementación de evaluaciones a gran escala.

TABLA 1
Características de Enem, Planea, Paebs y Exeims-BC

EVALUACIONES	BRASIL		MÉXICO	
	Nacional	Estatal	Nacional	Estatal
	Enem	Paebs	Planea	Exeims-BC
Año de implementación	1998	2000	2015	2012
Año de la matriz de referencia	2010	2022	2019	2020
Áreas de conocimiento (Brasil)	Ciencias naturales y sus tecnologías	Ciencias humanas	Lenguaje y comunicación	Ciencias experimentales
Campos disciplinarios (México)	Ciencias humanas y sus tecnologías	Ciencias de la naturaleza	Matemáticas	Humanidades y ciencias sociales
	Lenguajes, códigos y sus tecnologías	Lenguaje	Habilidades socioafectivas	Comunicación
	Matemáticas y su tecnología	Matemáticas		Matemáticas

(continua)

(continuação)

EVALUACIONES	BRASIL		MÉXICO	
	Nacional	Estatal	Nacional	Estatal
	Enem	Paebes	Planea	Exeims-BC
Componentes curriculares (Brasil) Subcampos (México)	Arte Biología Educación física Filosofía Física Geografía Historia Lengua extranjera Lengua portuguesa Literatura Matemáticas Química Sociología Tecnología de la información	Biología Física Geografía Historia Lengua portuguesa Matemáticas Química	Español Matemáticas	Biología Ciencias Humanas Ciencias sociales Español Física Formación cívica y ética Inglés Matemáticas Química
Metodología psicométrica utilizada	TRI	TRI y TCT	TRI y TCT	No identificado

Fuente: Elaboración de los autores con datos de la investigación.

En cuanto a los años de implementación de estas evaluaciones nacionales y estatales, es claro que en Brasil y el estado de Espírito Santo, este proceso ocurrió entre finales de los años 1990 y principios de los años 2000. En el contexto mexicano, a su vez, las evaluaciones a gran escala parecen haber tardado en implementarse en comparación con los exámenes brasileños.

Otro aspecto que se observa en la Tabla 1 corresponde a la comprensión de las áreas de conocimiento y componentes curriculares evaluados en Brasil, entendidos como campos y subcampos disciplinarios, respectivamente, en México. También se puede observar que en los campos disciplinarios del Exeims-BC están presentes las humanidades y ciencias sociales, que son consideradas como subcampos. Sin embargo, no fue posible identificar qué componentes curriculares constituyen estas áreas en México. Sin embargo, al analizar las competencias en las matrices de referencia del Exeims-BC, hay indicios de que humanidades estarían relacionadas con filosofía e historia, mientras que ciencias sociales incluyen los componentes curriculares geografía y sociología.

El análisis de las cuatro matrices de referencia tiene como objetivo identificar las metodologías psicométricas de análisis de datos utilizadas para generar resultados en estas evaluaciones. Por lo tanto, las metodologías psicométricas presentaron características distintas. En Enem, la TRI se utiliza para la construcción, análisis de ítems y cálculo de competencia, en línea con el carácter abstracto de los conocimientos evaluados en las matrices (Inep, 2022). Paebes, a su vez, además

de utilizar la TRI, también genera resultados a partir de la TCT, con porcentaje de aciertos por habilidad (descriptor) y total en el test. Para los estudiantes evaluados, también está la competencia alcanzada y una indicación del estándar de desempeño (Centro de Políticas Públicas e Avaliação da Educação [CAEd], s.d.).

Planea, al igual que Paebes, combina dos metodologías psicométricas de análisis de datos: TCT y TRI. La primera se utiliza para observar la distribución del desempeño de los estudiantes evaluados, comparando las competencias estimadas para responder al ítem con sus parámetros de dificultad. La segunda se refiere a realizar análisis comparativos entre subpoblaciones de estudiantes que, teniendo el mismo nivel de capacidad para responder a un ítem, no tienen la misma probabilidad de hacerlo correctamente (Inee, 2019).

En el caso de Exeims-BC, la falta de información precisa sobre la metodología psicométrica de análisis de datos subyacente resalta la importancia de cuestionar esta política de evaluación educativa (Gobierno del Estado de Baja California, 2012). La descripción de la clasificación de los resultados en niveles alcanzados proporciona evidencia de que la TRI es la metodología psicométrica que sustenta el análisis de los resultados del Exeims-BC. Según Jiménez (2014), la TCT, centrada en aspectos estadísticos como la confiabilidad y el error de medición, se basó en las ideas originales de Charles Spearman e implica la diferenciación entre puntaje verdadero y puntaje observado como resultado de la aplicación de una prueba. Sin embargo, a partir de la década de 1940, Stanley Smith Stevens clasificó los niveles de medición de las variables (nominales, ordinales, de intervalo y de razón), fortaleciendo la aplicación de métodos cuantitativos para medir características psicológicas en las personas y determinar medidas de variables sociales para el desarrollo de la estadística (Jiménez, 2014). La TRI, cuyas bases se desarrollaron en la segunda mitad del siglo XX, contó con contribuciones de las ideas de Louis Guttman, Frederic Lord y Georg Rasch. La centralidad de esta metodología psicométrica indica que el análisis de las respuestas en una prueba específica (evaluación a gran escala) se enfoca en sus componentes constitutivos (los ítems) en lugar del resultado global de la medición.

El avance tecnológico en las evaluaciones fundamentadas en esta metodología psicométrica ha contribuido a su efectividad, pero también ha impulsado el interés de las instituciones educativas por estandarizar contenidos y evaluar el desempeño de los estudiantes. En este sentido, surgió la necesidad de contar con criterios objetivos y eficientes para el ingreso a los diferentes niveles educativos. Este es un caso presente en Baja California para el acceso a la educación media superior, buscando asegurar que los nuevos estudiantes cuenten con los conocimientos o competencias mínimos indispensables para afrontar los retos académicos (Jiménez, 2014). En el

caso brasileño, el Enem también asume esa característica, al convertirse en una de las vías de ingreso a la educación superior, que, según Malusá et al. (2014), ayudó a estimular la reorganización del currículo de la escuela secundaria brasileña.

Centralidad de Enem y Paebes

Considerando las características culturales y políticas educativas de cada país y, al mismo tiempo, tomando como referencia que en el Paebes participan los estudiantes de secundaria del estado de Espírito Santo que también participan del Enem, y, de igual manera, los estudiantes de secundaria que participan en Planea también suelen tomar Exeims-BC, se optó por establecer análisis comparativos en función de cada país y su respectivo estado. En este sentido, se elaboró las figuras 1 y 2 en Iramuteq, resultantes de las matrices de referencia de los respectivos exámenes brasileños.

FIGURA 1

Nube de palabras de matriz de referencia de Enem



Fuente: Elaboración propia con ayuda del software Iramuteq.

FIGURA 2

Nube de palabras de matriz de referencia de Paebes



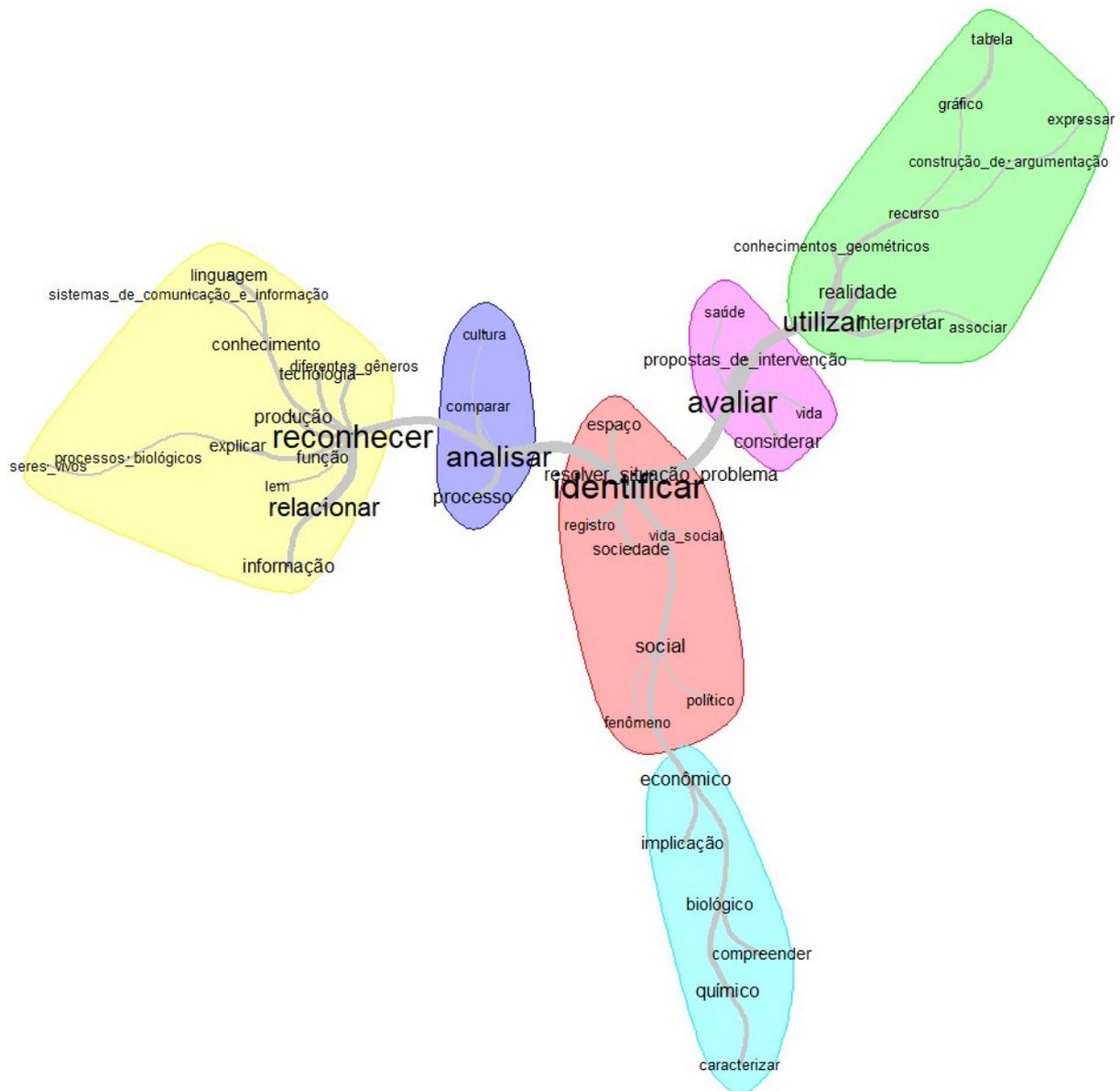
Fuente: Elaboración propia con ayuda del software Iramuteq.

Comparando las figuras 1 y 2, se desprende que la matriz de referencia de Enem, al abordar un mayor número de componentes curriculares, tiene más competencias para ser evaluadas y, en consecuencia, presentó mayores recurrencias de palabras (Figura 1). Por otro lado, la matriz de referencia de Paebes, al centrarse en un número más restringido de componentes curriculares, presentó menos recurrencia de palabras (Figura 2).

Aún con esta diferenciación, estas matrices de referencia también establecen aproximaciones en cuanto a las competencias evaluadas. Dado que existe poca diferencia en términos de competencias entre la evaluación nacional y la evaluación estatal, surge la pregunta sobre la necesidad de la existencia de estas dos evaluaciones educativas a gran escala en el mismo contexto, dado que las similitudes son significativas. Principalmente cuando se considera el Enem y su alcance nacional, al ser realizada por estudiantes de fin de secundaria en todo Brasil, es más diverso en cuanto al número de componentes curriculares y áreas de conocimiento evaluados en comparación con el Paebes, como se evidencia en las figuras 1 y 2.

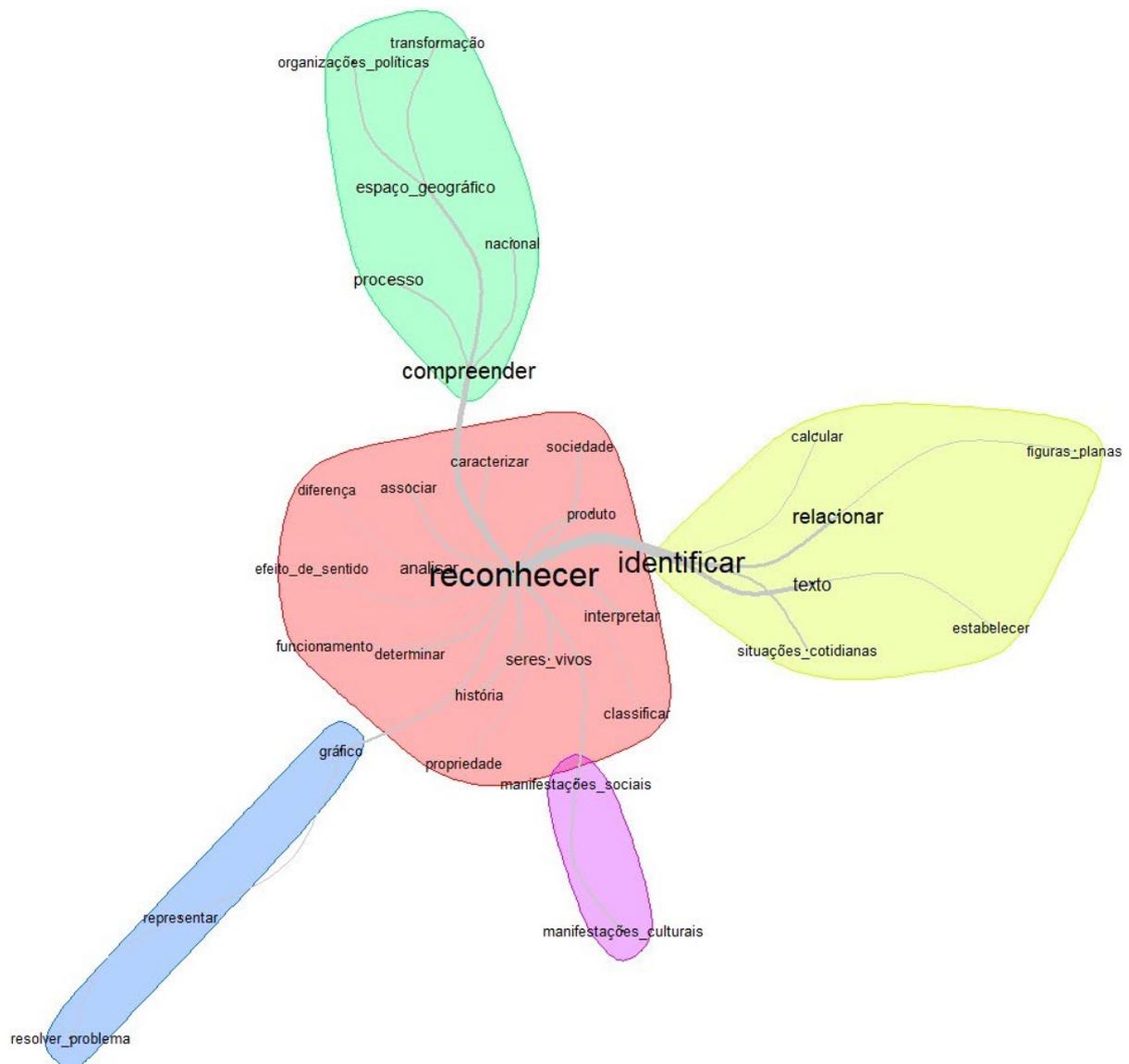
Para entender si las competencias de estas evaluaciones también son similares en cuanto a lo que se pretende alcanzar, se utilizó nuevamente Iramuteq para generar las figuras 3 y 4, que ordenan las palabras en clústeres (grupos de palabras) y ramas derivadas de las conexiones establecidas entre ellos según las fuentes. Los clústeres están dispuestos de tal manera que las palabras más recurrentes se ubican más hacia el centro de la figura, mientras que los grupos con palabras menos intensas se ubican periféricamente.

FIGURA 3
Similitud de palabras: Matriz de referencia de Enem



Fuente: Elaboración propia con ayuda del software Iramuteq.

FIGURA 4
Similitud de palabras: Matriz de referencia de Paebs



Fuente: Elaboración propia con ayuda del software Iramuteq.

En las figuras 3 y 4 se resaltan las palabras que se pueden comparar entre ambas matrices de referencia, ya que están presentes en las dos imágenes, y corresponden a los términos: *identificar*, *reconocer*, *comprender*, *analizar*, *relacionar* e *involucrar*. El tamaño destacado de ciertas palabras en las figuras se basa en la cantidad de repeticiones que tienen en las matrices de referencia. En la Figura 3, resultante de la matriz de referencia de Enem, la palabra *identificar*, representada por el clúster rojo, tiene una mayor expresividad en relación con las otras palabras. Según la Figura 4, *identificar* es la segunda palabra más destacada, demostrando ser aún más constante en la matriz de referencia de Paebs. Estas palabras aparecen en ejemplos como:

Identificar registros de práticas de grupos sociais no tempo e no espaço. . . .
Identificar etapas em processos de obtenção, transformação, utilização

ou reciclagem de recursos naturais, energéticos ou matérias-primas, considerando processos biológicos, químicos ou físicos neles envolvidos. (Matriz de referencia Enem [Inep, 2010, pp. 8-11]).

Identificar a relação entre o número de vértices, faces e/ou arestas de poliedros expressa em um problema. . . . Identificar o gráfico que representa uma situação descrita em um texto. . . . Identificar a presença de radiações em situações cotidianas (raios x, radiação solar, micro-ondas, entre outros). (Matriz de referencia Paebes [Sedu, 2022, pp. 1-4]).

La palabra central en la Figura 4 corresponde a *reconocer* y está orbitada por otros conjuntos de palabras (clústeres) con menor fuerza. El término *reconocer* se asocia en su propio clúster a diferentes contenidos y competencias desarrolladas por el propio alumno, es decir, no limitadas a la simple acción de reconocer algo. Asimismo, en la Figura 3, la palabra *reconocer* también está presente, pero como el segundo término más recurrente, estando asociada a otro conjunto de palabras menos expresivo de su grupo (naranja). Esto es evidente en los documentos de diferentes maneras:

Reconhecer as principais teorias sobre a origem e evolução dos seres vivos e suas características. . . . Reconhecer a importância do trabalho humano e suas formas de organização em diferentes contextos históricos. (Matriz de referencia Paebes [Sedu, 2022, pp. 2-3]).

Reconhecer a importância da produção cultural em LEM como representação da diversidade cultural e linguística. . . . Reconhecer em textos de diferentes gêneros, recursos verbais e não verbais utilizados com a finalidade de criar e mudar comportamentos e hábitos. . . . Reconhecer características ou propriedades de fenômenos ondulatórios ou fenômenos oscilatórios, relacionando-os a seus usos em diferentes contextos. (Matriz de referencia Enem [Inep, 2010, pp. 2, 4, 8]).

En la Figura 4, el clúster verde contiene la palabra *comprender* como la más destacada y está relacionado con cuestiones de naturaleza geopolítica. En la Figura 3, la acción de *comprender*, mostrada en el clúster azul claro, se centra en aspectos de naturaleza biológica, como se muestra en los fragmentos:

Compreender as transformações dos espaços geográficos como produtos-produtores das relações socioeconômicas e culturais de poder. . . . Compreender as múltiplas inter-relações dos processos de globalização e fragmentação do espaço geográfico. (Matriz de referencia Paebes [Sedu, 2022, pp. 1-3]).

Compreender o papel da evolução na produção de padrões, processos biológicos ou na organização taxonômica dos seres vivos. . . . Compreender fenômenos decorrentes da interação entre a radiação e a matéria em suas

manifestações em processos naturais ou processos tecnológicos, ou em suas implicações biológicas, sociais, econômicas ou ambientais. (Matriz de referencia Enem [Inep, 2010, pp. 9-10]).

La palabra *analizar* está presente en el clúster amarillo de la Figura 3 y en el clúster azul oscuro de la Figura 4. En la matriz de referencia de Enem se relaciona con factores cuantitativos y argumentativos, mientras que en Paebes se relaciona con cuestiones políticas y sociales:

Analisar informações envolvendo a variação de grandezas como recurso para a construção de argumentação. . . . Analisar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos. (Matriz de referencia Enem [Inep, 2010, p. 6]).

Analisar os impactos das transformações no meio ambiente pelas organizações políticas, econômicas, sociais e culturais. . . . Analisar o impacto das tecnologias na organização das sociedades. (Matriz de referencia Paebes [Sedu, 2022, p. 3]).

Otro término presente tanto en la Figura 3 como en la Figura 4 fue *relacionar*, apareciendo con mayor expresividad en el clúster amarillo de la Figura 3, refiriéndose principalmente al desarrollo de habilidades de lenguaje, comunicación e información, mientras que, en la Figura 4, aparece en el clúster amarillo, relacionado principalmente con características cuantitativas, espaciales y geográficas.

Relacionar informações geradas nos sistemas de comunicação e informação, considerando a função social desses sistemas. . . . Relacionar informações apresentadas em diferentes formas de linguagem e representação usadas nas ciências físicas, químicas ou biológicas, como texto discursivo, gráficos, tabelas, relações matemáticas ou linguagem simbólica. (Matriz de referencia Enem [Inep, 2010, pp. 2-9]).

Relacionar diferentes poliedros ou diferentes corpos redondos com suas planificações ou vistas. . . . Relacionar os diferentes tipos de atividade econômica às transformações do espaço geográfico. (Matriz de referencia Paebes [Sedu, 2022, pp. 1-3]).

Además de las palabras destacadas, que pueden compararse por ser frecuentes en ambas matrices de referencia, también hay palabras expresivas que no son comunes en las dos figuras, como por ejemplo *evaluar* (clúster rosa), *utilizar* (clúster verde claro) y *social* (clúster rojo), presentes en la Figura 3. La palabra *evaluar* se relaciona con aspectos biológicos y otros dos grupos que exponen aspectos cuantitativos, político-sociales y químicos. El término *uso* se refiere a contenidos cuantitativos, física y química.

Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos. . . . Avaliar métodos, processos ou procedimentos das ciências naturais que contribuam para diagnosticar ou solucionar problemas de ordem social, econômica ou ambiental. . . . Avaliar possibilidades de geração, uso ou transformação de energia em ambientes específicos, considerando implicações éticas, ambientais, sociais e/ou econômicas. (Matriz de referencia Enem [Inep, 2010, pp. 5, 9-10]).

En la Figura 4, el término *resolver problema* está asociado a la resolución de problemas y aspectos físicos, matemáticos y geométricos, mostrando que esta es una característica muy recurrente en la matriz de referencia de Paebes. En este sentido, se pueden identificar aproximaciones con la matriz de referencia de Enem, como se representa en la Figura 3, en la que *resolver una situación problemática* se asocia con características matemáticas y geométricas.

Resolver problema envolvendo o cálculo de perímetro de figuras planas. (Matriz de referencia Paebes [Sedu, 2022, p. 1]).

Resolver situação problema envolvendo conhecimentos numéricos. . . . Resolver situação problema que envolva conhecimentos geométricos de espaço e forma. (Matriz de referencia Enem [Inep, 2010, p. 5]).

Con base en el análisis de las figuras 3 y 4 y considerando las matrices de referencia de Enem y Paebes, se desprende que estas evaluaciones no tienen de todo la misma intención, es decir, presentan diferentes objetivos y competencias a evaluar en los estudiantes. Si, en un principio, pareció estar claro que tendrían muchas similitudes, excluyendo la necesidad de ambas evaluaciones, a partir de los análisis de similitudes fue posible percibir que se privilegia algunos aspectos en una que son menos centrados en la otra. Enem, por ejemplo, se preocupa más por evaluar el aprendizaje en el ámbito cognitivo, con inclinación, en algunos casos, hacia cuestiones actitudinales, que conciernen a la relación con la información, con los seres vivos, con el medio ambiente y con el mundo.

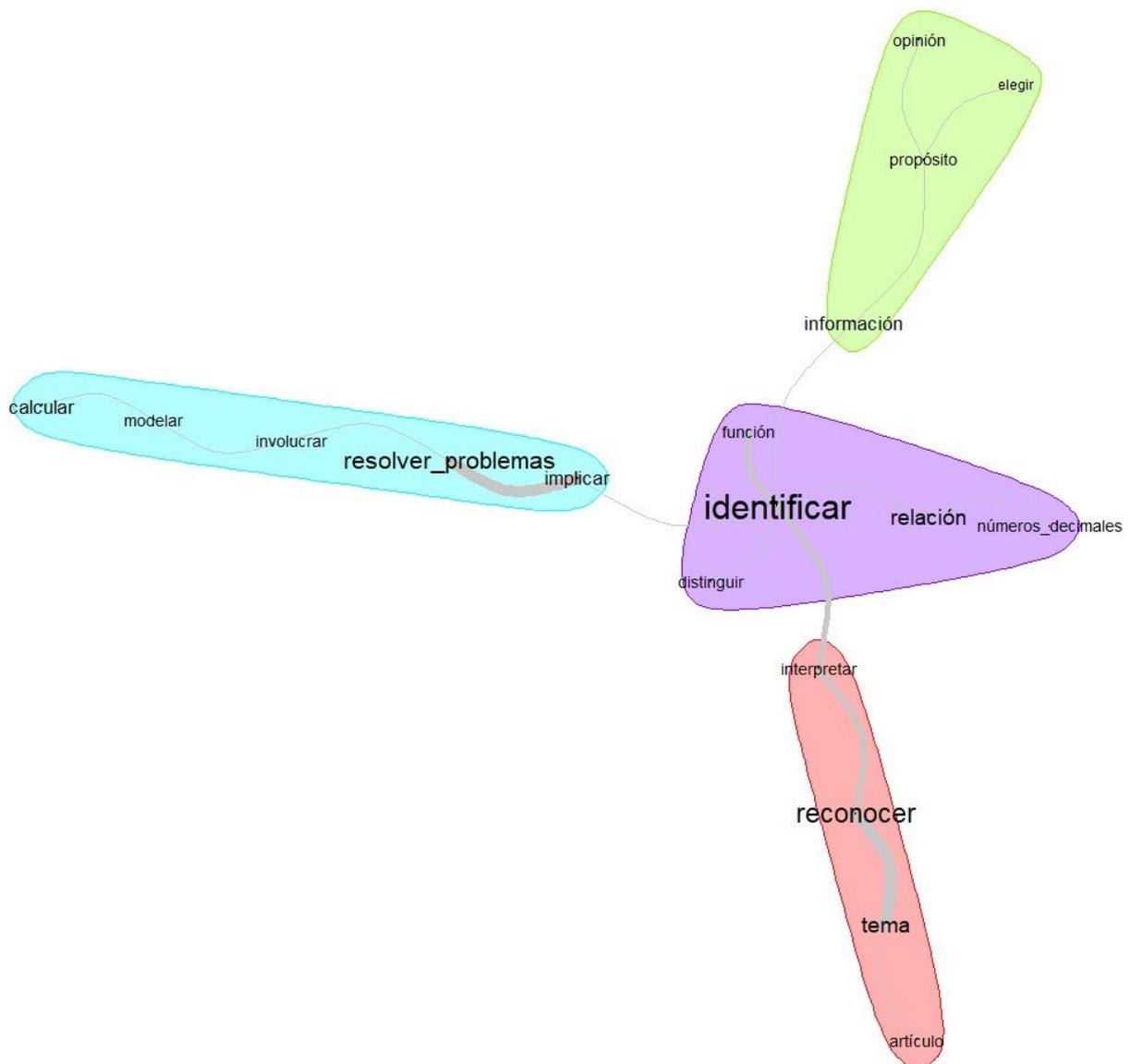
Centralidad de Planea y Exeims-BC

Centrándonos en el contexto de las evaluaciones educativas a gran escala en México, se produjeron las figuras 5 y 6, que corresponden, respectivamente, a las nubes de palabras obtenidas de las matrices de referencia Planea y Exeims-BC. Similar a lo que sucede en Brasil, donde los estudiantes que toman el Enem también toman el Paebes, en México los estudiantes residentes en Baja California participan tanto en Planea como en Exeims-BC, dada la cobertura de una nacional y otra local.

evaluación estatal involucra un mayor número de componentes curriculares y áreas evaluadas y, en consecuencia, una mayor variedad de competencias y habilidades estipuladas en su matriz de referencia. En este sentido, la Figura 5, correspondiente a Planea, indica menos palabras respecto a la Figura 6 de Exeims-BC.

Si bien existe una distinción, especialmente en lo que respecta al número de apariciones de palabras en las matrices de referencia de Planea y Exeims-BC, también se produjeron similitudes de palabras (figuras 7 y 8) en Iramuteq para identificar aproximaciones y distancias entre competencias evaluadas en ambas evaluaciones a gran escala.

FIGURA 7
Similitud de palabras: Matriz de referencia de Planea



Fuente: Elaboración propia con ayuda del software Iramuteq.

FIGURA 8
Similitud de palabras: Matriz de referencia de Exeims-BC



Fuente: Elaboración propia con ayuda del software Iramuteq.

Comparando las figuras 7 y 8, queda claro que ambas tienen resaltado el término *resolver problemas*. En el clúster azul de la Figura 7, *resolver problemas* se relaciona con cuestiones matemáticas, similar a lo que ocurre en la Figura 8.

Resolver problemas que impliquen calcular el área de polígonos regulares, o resolver problemas que impliquen conversiones entre medidas de volumen y de capacidad. (Manual técnico del Planea [Inee, 2019, p. 7]).

Resolver problemas aditivos con números fraccionarios y decimales. . . . Resolver problemas multiplicativos que impliquen el uso de expresiones algebraicas. (Tabla de especificaciones para Exeims [CoBachBC, 2020, pp. 3-6]).

También es posible observar, en la Figura 7, la presencia de palabras expresivas que no son comparables con la Figura 8. Se trata de la palabra *identificar*, presente en el clúster morado y asociada a la literatura, y *reconocer*, posicionado en el clúster rojo, refiriéndose principalmente a la construcción de argumentaciones, cálculos y conocimientos geométricos, como lo indican los fragmentos de la matriz de referencia de Planea.

Identificar la figura geométrica que sirve como modelo para recubrir un plano dado, figuras simétricas respecto a un eje oblicuo y la congruencia de triángulos, o resolver problemas que impliquen la relación entre un ángulo inscrito y ángulo central en una circunferencia, si ambos abarcan el mismo arco. . . . Identificar el argumento de mayor peso que utiliza el autor para apoyar su punto de vista; o distinguir dos puntos de vista. (Manual técnico del Planea [Inee, 2019, pp. 7-9]).

Reconocer la trama y el conflicto en un cuento e interpretar el lenguaje figurado de un poema. (Manual técnico del Planea [Inee, 2019, p. 11]).

En la Figura 8 aparecen en mayor proporción los términos *manejar técnicas*, *escribir*, *interpretar* y *analizar*. *Manejar técnicas*, evidente en el clúster rosa, no está relacionada con términos que indiquen una competencia específica. *Escribir* e *interpretar*, presentes en el clúster amarillo, están relacionados con aspectos lingüísticos, dialógicos y comunicativos. *Analizar*, presente en el clúster rosa, no está asociado con palabras de su propio clúster, sino con el conocimiento del lenguaje y la información.

Manejar técnicas eficientemente. (Tabla de especificaciones para Exeims [CoBachBC, 2020, p. 3]).

Escribir acuerdos y/o desacuerdos sobre un tema de estudio para intervenir en un debate. (Tabla de especificaciones para Exeims [CoBachBC, 2020, p. 14]).

Interpretar y ofrecer indicaciones para planear un paseo. (Tabla de especificaciones para Exeims [CoBachBC, 2020, p. 13]).

Analizar la información y emplear el lenguaje para la toma de decisiones (Tabla de especificaciones para Exeims [CoBachBC, 2020, p. 10]).

Al comparar las figuras 5 a 8, se identifica que, a diferencia de lo evidenciado en las matrices de referencia de Enem y Paebes, otras competencias parecen ganar protagonismo, especialmente en lo que respecta a Exeims-BC, en el que habilidades relacionadas con aspectos sociales son más recurrentes. También es posible notar menos similitud entre la evaluación educativa nacional y estatal a gran escala en México, ya que el número de palabras es significativamente menor en el examen nacional (Planea) en comparación con el examen estatal (Exeims-BC).

En general, los resultados demuestran el predominio de competencias a alcanzar y demostrar a través de estas políticas de evaluación, centrándose en dos áreas básicas, a saber, matemáticas y lengua materna. Por otro lado, fue evidente la presencia, aunque tímida, de competencias evaluadas que involucran aspectos procedimentales y actitudinales. Esto demuestra la necesidad de continuar investigando y discutiendo las implicaciones políticas de las evaluaciones educativas a gran escala en contextos macro- (internacional), meso- (nacional) y microsocio (estatal).

DISCUSIONES Y CONCLUSIONES

Este artículo buscó realizar un análisis comparativo de las matrices de referencia de Enem (Brasil) y Planea (México), junto con sus evaluaciones estatales, Paebs y Exeims-BC, respectivamente. El uso del análisis documental y del *software* Iramuteq contribuyó a contextualizar y comparar las matrices de referencia, brindando una visión general de la centralidad de estas evaluaciones. El análisis reveló características específicas de cada contexto, identificando elementos comparables e incomparables entre las políticas de evaluación.

Al iniciar la discusión, se destaca que Enem, en el escenario brasileño, cubre una mayor variedad de componentes curriculares y competencias que Paebs (Espírito Santo), que se enfoca en un conjunto más restringido de materias. Esto plantea dudas sobre la necesidad de mantener ambas evaluaciones, considerando que Enem, debido a su alcance nacional, ya evalúa las competencias cubiertas en Paebs. En México, Exeims-BC tiene mayor amplitud que Planea, lo que sugiere variaciones en el abordaje de las competencias (Mattos Santos et al., 2023).

Es esencial comprender las motivaciones que guiaron las decisiones de los gobiernos nacionales y estatales a lo largo de los años al implementar estas evaluaciones, que funcionan como mecanismos de control estatal para evaluar la calidad de la educación nacional y estatal, con el objetivo de obtener datos con mayor frecuencia y promover reformas en las políticas educativas (Cunha & Müller, 2018).

Menegão (2017) destaca que los sistemas de evaluación –incluidos Enem, Paebs, Planea y Exeims-BC– ejercen un control curricular que favorece un conjunto específico de conocimientos, limitando la perspectiva educativa. Esta restricción ha preocupado a los estudiosos de las políticas de evaluación, principalmente por el carácter neoliberal evidente en estos enfoques (Gentili & Silva, 1994; Mons, 2009; Freitas et al., 2012; Monarca, 2015; Dubet, 2016; Parra & Matus, 2016; Carrasco & Gunter, 2019; Ozga, 2020). Estas políticas están moldeadas por el nuevo orden internacional, que determina los temas prioritarios de evaluación (Verger et al.,

2017, 2019). A pesar de las menciones de otras materias en las fuentes, sorprende observar que componentes curriculares como sociología, filosofía, artes y educación física aún no han sido incorporadas a todas las evaluaciones analizadas (Mattos Santos et al., 2024) a excepción de Enem.

Se puede observar que tres evaluaciones tienen como metodología psicométrica de análisis de datos la TRI, evidenciando una acción de introducir medidas adicionales y herramientas más sofisticadas (en comparación con la TCT), lo que es la consecuencia lógica de la creación de organizaciones especializadas (normalmente agencias de evaluación más sofisticadas) cuya principal responsabilidad es supervisar y utilizar un volumen cada vez mayor de datos recopilados (Verger et al., 2019).

Para Martínez (2008), las evaluaciones a gran escala han tenido como principal objetivo identificar el nivel de aprendizaje alcanzado por un país, región o distrito. En el caso concreto de este estudio, por un lado, se evidenció la presencia de la TRI en al menos tres evaluaciones. Según Leenen (2014), incluso con el mayor uso de la TRI a lo largo de las décadas y siendo considerada una alternativa viable entre los expertos en metodologías psicométricas, el principal enfoque en contextos aplicados para analizar los resultados de evaluaciones a gran escala sigue siendo la TCT. Este hecho se revela cuando se identifica que, de las cuatro evaluaciones analizadas en este estudio, dos (Planea y Paebs) combinan metodologías psicométricas de análisis de datos (TCT y TRI), lo que resalta, además de la aproximación de la evaluación nacional de México con la evaluación estatal de Espírito Santo (Brasil), que esto no es específico de un contexto determinado.

Al analizar la centralidad de las competencias evaluadas según las matrices de referencia, se observan similitudes en las evaluaciones nacionales y estatales. Esto se evidencia en palabras como *identificar*, *reconocer* y *comprender*, que son recurrentes en todas las matrices, resaltando expectativas de aprendizaje relacionadas con el ámbito cognitivo.

Respecto a la intencionalidad de las evaluaciones educativas a gran escala según sus matrices de referencia, el presente estudio sugiere que, a pesar de algunas similitudes, las evaluaciones tienen diferentes competencias para ser consideradas en exámenes estandarizados. Mientras que, en Brasil, Enem se preocupa más por evaluar el aprendizaje en el ámbito cognitivo, con inclinación hacia cuestiones actitudinales relacionadas con el medio ambiente, los seres vivos y el mundo, centrándose en habilidades más específicas. En México, aunque los resultados demostraron que expresiones como *resolver problemas* destacaron en las matrices de referencia, la naturaleza específica de los problemas difiere entre Planea y Exeims-BC. Mientras que en el primer caso la *resolución de problemas* se relaciona principalmente con el componente curricular de matemáticas, el énfasis

del Exeims-BC en proponer situaciones de *resolución de problemas* está más dirigido al área de las ciencias sociales.

En este sentido, se observa que la centralidad adquirida por evaluaciones de gran escala, como Enem, Paebes, Planea y Exeims-BC, sigue una lógica interna ampliamente difundida en pruebas educativas estandarizadas en otros países. A estos exámenes se les atribuyen funciones regulatorias, especialmente cuando se convierten en dispositivos con implicaciones significativas para la vida de estudiantes, docentes y unidades escolares, dado su fundamento en el modelo de rendición de cuentas (Miras & Solé, 1996; Linn, 2008; Baker, 2009; Álvarez-López & Matarranz García, 2020). En consecuencia, los resultados de estos exámenes suelen estar relacionados con: a) la certificación de los estudiantes y el acceso a determinados cursos de formación técnica superior (en el caso de Exeims-BC) o cursos de educación superior, como en el caso de Enem en Brasil y Planea en México, según los resultados de la evaluación; b) incentivos y sanciones a los docentes, muchas veces en forma de beneficios económicos, como ocurre en el estado de Espírito Santo a partir de los resultados de Paebes (Lopes et al., 2024).

La falta de información específica sobre la metodología psicométrica de análisis de datos que sustenta Exeims-BC resalta la importancia de la transparencia en la información relacionada con las políticas de evaluación educativa, destacando la necesidad de divulgar las metodologías psicométricas que sustentan estas prácticas.

El análisis comparativo de las matrices de referencia de evaluaciones educativas a gran escala en Brasil y México resalta posibles influencias de las evaluaciones nacionales en las evaluaciones estatales, al mismo tiempo que son evidentes las especificidades que reflejan contextos culturales y educativos únicos. Dada la importante similitud en las competencias evaluadas, que indica posibles redundancias en las competencias evaluadas, surge como una reflexión, en línea con Uczak (2014), la necesidad de dos evaluaciones educativas a gran escala en un mismo contexto educativo para un mismo grupo de estudiantes, en este caso, de secundaria.

El estudio, de acuerdo con las recomendaciones de Bauer et al. (2015) y Steiner-Khamsi (2014, 2016), resalta la importancia de considerar las bases (metodologías psicométricas) en la construcción y análisis de matrices de referencia, promoviendo la posibilidad de comprender las implicaciones de las evaluaciones educativas a gran escala en los sistemas nacionales y estatales. También es evidente que estas evaluaciones establecen una relación con las políticas educativas globales (Verger et al., 2017, 2019), pues existen alineamientos entre ellas y con evaluaciones internacionales como Pisa, por ejemplo.

En el contexto mexicano, la investigación revela un escenario opuesto, en el que el Exeims-BC (evaluación estatal) cubre una mayor variedad de componentes curriculares y competencias evaluadas en comparación con Planea. El análisis utilizando las figuras generadas en Iramuteq indicó que las competencias enfatizadas en las evaluaciones educativas mexicanas a gran escala tienen diferencias más pronunciadas que las observadas en las evaluaciones educativas brasileñas a gran escala.

En general, ambos contextos (brasileño y mexicano) presentan similitudes en el enfoque de las evaluaciones educativas, pero también revelan matices específicos de cada sistema. Además, los resultados resaltan la complejidad de las políticas de evaluación educativa en diferentes contextos nacionales y estatales. Aunque se identifican algunos estándares de competencia, las diferencias entre las evaluaciones educativas a gran escala nacionales y estatales en los dos países demuestran la necesidad de un análisis cuidadoso al diseñar políticas educativas y evaluaciones a gran escala. La investigación enfatiza la importancia de evaluar no sólo el conocimiento cognitivo, sino también las habilidades procedimentales y actitudinales, para obtener una comprensión más integral de los procesos de aprendizaje de los estudiantes.

Sin embargo, es fundamental cuestionar la motivación detrás de esta multiplicación de evaluaciones educativas a gran escala y evaluar si realmente sirven a un currículo instrumental y pretenden responsabilizar a la escuela y al profesorado por el desempeño de los estudiantes (Álvarez-López & Matarranz García, 2020). Este modelo educativo, dirigido a países en desarrollo e influenciado por organismos internacionales como el Banco Mundial y la Unesco, prioriza la preparación inmediata para el mercado laboral (Libâneo & Freitas, 2018). En consecuencia, los conocimientos y las competencias esperadas en las evaluaciones siguen esta misma lógica, ya sea a escala local (estatal), nacional e incluso internacional.

Al tratarse de una investigación con enfoque cualitativo, se destacan algunas limitaciones del presente estudio. Principalmente en lo que respecta a las evaluaciones estatales, dado que otros estados presentan sus propias políticas de evaluación, lo que puede resultar en otros hallazgos. Por lo tanto, es necesario considerar que los resultados representan sólo una porción del universo de evaluaciones a gran escala presentes en los sistemas educativos de Brasil y México, lo que no permite hacer generalizaciones para todo el contexto brasileño y mexicano, sino más bien constituir un análisis específico.

Se concluye que, en las cuatro evaluaciones educativas a gran escala, las competencias requeridas en las matrices de referencia abarcan principalmente conocimientos en las áreas de matemáticas y lengua materna. Esto refuerza la necesidad de desarrollar nuevos estudios que investiguen cómo estas matrices han

guiado las prácticas pedagógicas de los profesionales que actúan en la educación básica en los contextos en los que se insertan estas políticas evaluativas. Finalmente, este estudio destaca la continua relevancia de investigar y discutir las implicaciones de las políticas de evaluaciones a gran escala en diferentes niveles, desde contextos macro a micro, con el fin de mejorar el sistema educativo y brindar una educación de calidad para todos los estudiantes.

AGRADECIMIENTOS

Los autores agradecen a Fundação de Amparo à Pesquisa e Inovação do Espírito Santo por su apoyo y promoción de la investigación a través de la beca de Iniciación Científica del Proyecto Fapes n. 11367/2021, y para la beca Postdoctoral de Profix Circular n. 15/2022, Protocolo: 51005.788.17880.

REFERENCIAS

- Álvarez-López, G., & Matarranz García, M. (2020). Calidad y evaluación como tendencias globales en política educativa: Estudio comparado de agencias nacionales de evaluación en educación obligatoria en Europa. *Revista Complutense de Educación*, 31(1), 83-93. <https://doi.org/10.5209/rced.61865>
- Baird, J., Johnson, S., Hopfenbeck, T., Isaacs, T., Sprague, T., Stobart, G., & Yu, G. (2016). On the supranational spell of PISA in policy. *Educational Research*, 58(2), 121-138. <https://doi.org/10.1080/00131881.2016.1165410>
- Baker, E. (2009). Consideraciones de validez prioritaria para la evaluación formativa y de rendición de cuentas. *Revista de Educación*, (348), 91-110.
- Bauer, A., Alavarse, O. M., & Oliveira, R. P. (2015). Avaliação em larga escala: Uma sistematização do debate. *Educação e Pesquisa*, 41(especial), 1367-1382. <https://doi.org/10.1590/S1517-9702201508144607>
- Bazán, J. L. (2018). Psicometria e avaliação por testes: Um marco metodológico. In J. C. Rothen, & A. C. M. Santana (Eds.), *Avaliação da educação: Referências para uma primeira conversa* (pp. 139-156). EdUFSCar.
- Bloch, M. (1998). Para uma história comparada das sociedades europeias. In M. Bloch, *História e historiadores* (pp. 119-150). Teorema.
- Bloch, M. (2001). *Apologia da história ou o ofício do historiador*. Jorge Zahar.
- Camargo, B. V., & Justo, A. M. (2013). Iramuteq: Um software gratuito para análise de dados textuais. *Temas em Psicologia*, 21(2), 513-518. <http://dx.doi.org/10.9788/TP2013.2-16>
- Carrasco, A., & Gunter, H. M. (2019). The “private” in the privatisation of schools: The case of Chile. *Educational Review*, 71(1), 67-80. <https://doi.org/10.1080/00131911.2019.1522035>
- Casari, M. (2016). Programa Internacional de Avaliação de Estudantes (PISA): A concepção de letramento e o estado da arte no Brasil. *Trama*, 12(27), 84-109. <https://e-revista.unioeste.br/index.php/trama/article/view/14458>
- Centro de Políticas Públicas e Avaliação da Educação (CAEd). (s.d.). *Resultados – Avaliação e monitoramento Espírito Santo*. CAEd. <https://avaliacaoemontoramentoespiritosanto.caeddigital.net/#!/resultados>

- Chartier, J. F., & Meunier, J. G. (2011). Text mining methods for social representation analysis in Large Corpora. *Papers on Social Representations*, 20(37), 1-47.
- Colégio de Bachilleres del Estado de Baja California (CoBachBC). (2020). *Tabla de especificaciones para EXEIMS*. CoBachBC.
- Creswell, J. W., & Clark, V. L. P. (2013). *Pesquisa de métodos mistos*. Penso.
- Cunha, E. C. S., & Müller, E. R. (2018). Avaliações em larga escala: Uma tentativa de controle, regulação, captura e padronização do cotidiano escolar. *Cadernos da Fucamp*, 17(29), 143-163. <https://revistas.fucamp.edu.br/index.php/cadernos/article/view/1317>
- Dubet, F. (2016). *Ce qui nous unit: Discriminations, égalité et reconnaissance*. Seuil.
- Elías, R. (2017). Los programas internacionales de evaluación de logros académicos y su influencia en las políticas educativas en América Latina. *Población y Desarrollo*, 23(45), 74-82. [https://doi.org/10.18004/pdfce/2076-054x/2017.023\(45\)074-082](https://doi.org/10.18004/pdfce/2076-054x/2017.023(45)074-082)
- Ferreira, J., Salgado, A. P., Junior, Novi, J. C., Miura, I. K., & Diogo, D. O. (2017). Estudo exploratório sobre eficiência nas escolas municipais paulistas: Melhores práticas e desempenho no Ideb. *Meta: Avaliação*, 9(25), 32-64. <http://dx.doi.org/10.22347/2175-2753v9i25.1240>
- Fletcher, P. R. (2010). *Da teoria clássica dos testes para os modelos de resposta ao item*. Escola Nacional de Ciências Estatísticas.
- Freitas, L. C., Malvasi, M. M. S., Sordi, M. R. L., Mendes, G. S. C. V., & Almeida, L. C. (2012). *Avaliação e políticas públicas educacionais: Ensaio contrarregulatórios em debate*. Edições Leitura Crítica.
- Gentili, P. A. A., & Silva, T. T. da. (Orgs.). (1994). *Neoliberalismo, qualidade total e educação*. Vozes.
- Gobierno del Estado de Baja California. (2012). *La evaluación educativa en Baja California: Construcción de una cultura*. Gobierno del Estado de Baja California. http://www.educacionbc.edu.mx/departamentos/evaluacion/LibroEvaluaciones/ESTADO_DEL_ARTE_Evaluacion_TODO.pdf
- González, I. B., & Bosco, E. B. (2006). *Políticas y sistemas de evaluación educativa en México: Avances, logros y desafíos*. Instituto Nacional para la Evaluación de la Educación.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2010). *Matriz de referência Enem*. Inep.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2022). *Parâmetros dos itens são divulgados de forma inédita*. Inep. <https://www.gov.br/inep/pt-br/assuntos/noticias/enem/parametros-dos-itens-sao-divulgados-de-forma-inedita>
- Instituto Nacional para la Evaluación de la Educación (Inee). (2019). *Manual técnico del Plan Nacional para la Evaluación de los Aprendizajes*. Inee.
- Jiménez, J. A. M. (2014). Evaluación a gran escala: Su desarrollo e inconvenientes en su modelo teórico. In *Actas del Congreso Internacional de Educación* (p. 1-11). Universidad Autónoma de Tlaxcala. https://www.researchgate.net/publication/312533589_Evaluacion_a_gran_escala_su_desarrollo_e_inconvenientes_en_su_modelo_teorico
- Leenen, I. (2014). Virtudes y limitaciones de la teoría de respuesta al ítem para la evaluación educativa en las ciencias médicas. *Investigación en Educación Médica*, 3(9), 40-55. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-50572014000100007
- Lei n. 9.394, de 20 de dezembro de 1996. (1996). Estabelece as diretrizes e bases da educação nacional. Brasília, DF. https://www.planalto.gov.br/ccivil_03/leis/19394.htm
- Ley General de Educación, del 13 de julio de 1993. (1993). Ciudad de México.

- Libâneo, J. C., & Freitas, R. M. da M. (Orgs.). (2018). *Políticas educacionais neoliberais e escola pública: Uma qualidade restrita de educação escolar*. Espaço Acadêmico.
- Linn, R. L. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, 40(6), 699-711. <https://doi.org/10.1080/00220270802105729>
- Lopes, I. P., Stieg, R., & Santos, W. (2024). Evaluaciones educativas a gran escala en los estados de Espírito Santo y Baja California: Especificidades y aproximaciones entre Paebs y Exeims-BC. *REXE- Revista de Estudios y Experiencias en Educación*, 23(52), 124-141. <https://doi.org/10.21703/rexe.v23i52.2349>
- Malusá, S., Ordonez, L. L., & Ribeiro, E. (2014). Enem: Pontos positivos para a educação brasileira. *Revista Educação e Políticas em Debate*, 3(2), 358-382. <https://doi.org/10.14393/REPOD-v3n2a2014-30284>
- Martínez, F. (2008). *Evaluación a gran escala y evaluación en el aula: Los usos de Excale para la mejora educativa*. Inee.
- Mattos Santos, P., Stieg, R., Barcelos, M., Ferreira, A., Neto, & Santos, W. (2023). La educación física como componente curricular evaluado en exámenes estandarizados nacionales en América y Europa. *Calidad en la Educación*, (59), 158-190. <https://doi.org/10.31619/caledu.n59.1318>
- Mattos Santos, P., Stieg, R., Barcelos, M., & Santos, W. dos. (2024). Evaluaciones nacionales a gran escala y acceso a la educación superior: perspectivas en países de América y Europa. *Contextos: Estudios de Humanidades y Ciencias Sociales*, (54), 1-25. <https://revistas.umce.cl/index.php/contextos/article/view/2660>
- Menegão, R. de C. S. G. (2017). Os impactos da avaliação em larga escala nos currículos escolares. *Práxis Educativa*, 11(3), 641-656. <https://doi.org/10.5212/PraxEduc.v11i3.0007>
- Miras, M., & Solé, I. (1996). A evolução da aprendizagem e a evolução do processo de ensino e aprendizagem. In C. Coll, J. Palacios, & A. Marchesi (Eds.), *Desenvolvimento psicológico e educação: Psicologia da educação* (pp. 201-219). Artes Médicas.
- Monarca, H. (2015). *Evaluaciones externas: Mecanismos para la configuración de representaciones y prácticas en educación*. Miño y Dávila.
- Mons, N. (2009). Effets théoriques et réels des politiques d'évaluation standardisée. *Revue Française de Pédagogie*, (169), 99-140. <https://doi.org/10.4000/rfp.1531>
- Ozga, J. (2020). The politics of accountability. *Journal of Educational Change*, 21, 19-35. <https://doi.org/10.1007/s10833-019-09354-2>
- Parra, V., & Matus, G. (2016). Usos de datos y mejora escolar: Una aproximación a los sentidos y prácticas educativas subyacentes a los procesos de toma de decisiones. *Calidad en la Educación*, (45), 207-250. <https://doi.org/10.31619/caledu.n45.25>
- Pasquali, L. (2017). *Psicometria: Teoria dos testes na psicologia e na educação*. Vozes.
- Pasquali, L., & Primi, R. (2003). Fundamentos da teoria da resposta ao item: TRI. *Avaliação Psicológica*, 2(2), 99-110. https://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712003000200002
- Perrenoud, P. (1999). *Construir competências desde a escola*. Artmed.
- Phillips, D., & Ochs, K. (2003). Processes of policy borrowing in education: Some explanatory and analytical devices. *Comparative Education*, 39(4), 451-461. <https://doi.org/10.1080/0305006032000162020>
- Rocha, L. D. U., Júnior. (2018). *Calibração e validação da versão brasileira do banco de itens relações sociais do Patient Related Outcomes Measures Information System (PROMIS®) pediátrico* [Tese de doutorado, Universidade Federal de Uberlândia]. Repositório Institucional – Universidade Federal de Uberlândia. <https://repositorio.ufu.br/handle/123456789/24422>

- Secretaria de Estado da Educação do Espírito Santo (Sedu). (2022). *Matriz de referência Paebs*. Sedu.
- Sellar, S., & Lingard, B. (2013). The OECD and the expansion of PISA: New global modes of governance in education. *British Educational Research Journal*, 40(6), 917-936. <https://doi.org/10.1002/berj.3120>
- Steiner-Khamsi, G. (2014). Cross-national policy borrowing: Understanding reception and translation. *Asia Pacific Journal of Education*, 34(2), 153-167. <https://doi.org/10.1080/02188791.2013.875649>
- Steiner-Khamsi, G. (2016). New directions in policy borrowing research. *Asia Pacific Education Review*, 17, 381-390. <https://doi.org/10.1007/s12564-016-9442-9>
- Uczak, L. H. (2014). *O Preal e as políticas de avaliação educacional para a América Latina* [Tese de doutorado, Universidade Federal do Rio Grande do Sul]. Lume Repositório Digital. <http://hdl.handle.net/10183/94732>
- Verger, A., Fontdevila, C., & Parcerisa, L. (2019). Reforming governance through policy instruments: How and to what extent standards, tests and accountability in education spread worldwide. *Discourse: Studies in the Cultural Politics of Education*, 40(2), 248-270. <https://doi.org/10.1080/01596306.2019.1569882>
- Verger, A., Moschetti, M., & Fontdevila, C. (2017). *La privatización educativa en América Latina: Una cartografía de políticas, tendencias y trayectorias*. Internacional de la Educación.
- Verger, A., Novelli, M., & Kosar, H. (2012). *Global education policy and international development: New agendas, issues and policies*. Bloomsbury Academic.
- Volante, L. (2015). The impact of PISA on education governance: Some insights from highly reactive policy contexts. *International Studies in Educational Administration*, 43(2), 103-117.
- Werle, F. O. C. (2011). Políticas de avaliação em larga escala na educação básica: Do controle de resultados à intervenção nos processos de operacionalização do ensino. *Ensaio: Avaliação e Políticas Públicas em Educação*, 19(73), 769-792. <https://doi.org/10.1590/S0104-40362011000500003>
- Zabala, A. (1998). *A prática educativa: Como ensinar*. Artes Médicas.

NOTA: Las contribuciones de cada autor en la elaboración del artículo fueron las siguientes: Mateus Gobbi dos Santos –recolección de datos, organización e interpretación de datos, redacción del manuscrito original; Ronildo Stieg – organización e interpretación de datos, revisión, edición y supervisión de redacción de datos; Berenice Morales González –recolección de datos y revisión de escrita; Wagner dos Santos –interpretación de datos, revisión y supervisión.