

Piedra Rosetta documento de política Establecimiento de una concordancia entre las evaluaciones regionales (ERCE y PASEC) e internacionales (TIMSS/PIRLS)



UNESCO

La Constitución de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) fue aprobada por 20 países en la Conferencia de Londres, en noviembre de 1945, y entró en vigor el 4 de noviembre de 1946.

El principal objetivo de la UNESCO es contribuir a la paz y la seguridad en el mundo promoviendo, mediante la educación, la ciencia, la cultura y la comunicación, la colaboración entre las naciones, a fin de asegurar el respeto universal de la justicia, el estado de derecho, los derechos humanos y las libertades fundamentales que la Carta de las Naciones Unidas reconoce a todos los pueblos del mundo sin distinción de raza, sexo, idioma o religión.

Para cumplir este mandato, la UNESCO desempeña cinco funciones principales: 1) estudios prospectivos sobre la educación, la ciencia, la cultura y la comunicación para el mundo del futuro; 2) el fomento, la transferencia y el intercambio del conocimiento mediante actividades de investigación, formación y educación; 3) acciones normativas, para la preparación y aprobación de instrumentos internos y recomendaciones estatutarias; 4) conocimientos especializados que se transmiten a los Estados Miembros mediante cooperación técnica para que elaboren sus políticas y proyectos de desarrollo; y 5) el intercambio de información especializada.

Instituto de Estadística de la UNESCO

El Instituto de Estadística de la UNESCO (UIS) es la oficina de estadística de la UNESCO y es el depósito de la ONU en materia de estadísticas mundiales en los campos de la educación, la ciencia y la tecnología la cultura y la comunicación. El UIS fue fundado en 1999. Se creó con el fin de mejorar el programa de estadística de la UNESCO, así como para desarrollar y suministrar estadísticas exactas, oportunas y políticamente relevantes, requeridas en un contexto actual cada vez más complejo y rápidamente cambiante.

Publicado en 2022 por:

Instituto de Estadística de la UNESCO

C.P 250 Succursale H

Montréal, Québec H3G 2K8

Canada

Correo electrónico: uis.publications@unesco.org

<http://www.uis.unesco.org>

Ref: UIS/2022/LO/TD/15

ISBN: 978-92-9189-300-3

© UNESCO-UIS 2022



Esta publicación está disponible en acceso abierto bajo la licencia Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). Al utilizar el contenido de la presente publicación, los usuarios aceptan las condiciones de uso del Repositorio de UNESCO de acceso abierto (www.unesco.org/open-access/terms-use-ccbysa-sp). Los términos empleados en esta publicación y la presentación de los datos que en ella aparecen no implican toma alguna de posición de parte de la UNESCO en cuanto al estatuto jurídico de los países, territorios, ciudades o regiones ni respecto de sus autoridades, fronteras o límites. Las ideas y opiniones expresadas en esta publicación son las de los autores; no son necesariamente las de UNESCO y no comprometen a la Organización.

Diseño y presentación: Freire Creative

Diseño de cubierta: büro Svenja

Agradecimientos

El Informe de Análisis de Rosetta Stone fue un proyecto colaborativo del Instituto de Estadística de la UNESCO (UIS). La Asociación Internacional para la Evaluación del Logro Educativo (IEA por sus siglas en inglés) fue el socio técnico de este proyecto junto con el Centro Internacional de Estudios TIMSS & PIRLS, Lynch School of Education and Human Development, Boston College. El apoyo técnico y de implementación fue proporcionado por CONFEMEN y ERCE.

Andrés Sandoval Hernández es el autor de este informe. Luis Crouch (RTI), Silvia Montoya (UIS), Dirk Hastedt (IEA) y Oliver Neuschmidt (IEA) han actuado como revisores.

Este informe se basa en una investigación financiada por la Fundación Bill y Melinda Gates. Los hallazgos y conclusiones contenidos en este informe son los de los autores y no reflejan necesariamente las posiciones o políticas de la Fundación Bill y Melinda Gates.

Este informe contiene extractos de los informes técnicos de Rosetta Stone proporcionados por el Centro Internacional de Estudios TIMSS & PIRLS sobre el desarrollo de las Concordancias TIMSS y PIRLS para ERCE y PASEC. Los informes técnicos están disponibles en <https://timssandpirls.bc.edu/Rosetta-Stone-Reports/index.html> y más preguntas sobre estos deben dirigirse al Centro Internacional de Estudios TIMSS & PIRLS: <https://timssandpirls.bc.edu/>

ÍNDICE

Lista de tablas	4
Introducción	5
Métodos	7
A. El diseño complejo de las ILSA	7
B. El Método de la Piedra de Rosetta	8
Las tablas de concordancia	11
Proporción de estudiantes que logran el ODS 4.1.1	14
Mensajes clave	19
Lecturas sugeridas	21
Lista de siglas	22

LISTA DE TABLAS

TABLA 1. Tabla de concordancia para matemáticas ERCE y PASEC.12

TABLA 2. Tabla de concordancia para comprensión lectora
ERCE y PASEC.13

TABLA 3. Definición de los puntos de referencia ERCE, PASEC
y TIMSS/PIRLS que sirven como MPLs para Grado 6 (ERCE y
PASEC) y Grado 4 (TIMSS/PIRLS)15

TABLA 4. Proporción de estudiantes en países ERCE y PASEC
que logran el ODS 4.1.1 en Matemáticas17

TABLA 5. Proporción de estudiantes en países ERCE y PASEC
que logran el ODS 4.1.1 en comprensión lectora18

INTRODUCCIÓN¹

Cuando quedan menos de diez años para alcanzar los 17 Objetivos de Desarrollo Sostenible (ODS) establecidos en la Agenda 2030, nuestros líderes mundiales han declarado una Década de Acción y ejecución para el desarrollo sostenible. La Década de Acción llama a acelerar las soluciones sostenibles a todos los mayores desafíos del mundo, desde la pobreza y el género hasta el cambio climático, la desigualdad y la mejora de la calidad de la educación para todos.

Este documento se centra en el Objetivo de Desarrollo Sustentable (ODS) 4, en particular el Indicador 4.1.1. El ODS 4 establece que para 2030 debemos “*garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje a lo largo de toda la vida para todos*”, y el Indicador 4.1.1 operacionaliza este objetivo en la demanda de “*garantizar que todas las niñas y niños completen educación primaria y secundaria gratuita, equitativa y de calidad que conduzca a resultados de aprendizaje pertinentes y efectivos*”.² La comunidad internacional acordó medir el progreso de esta meta como el porcentaje de niños y jóvenes que logran al menos un nivel mínimo de competencia en competencia lectora y matemáticas en tres puntos de la estructura de grados, y por sexo: (a) en los grados 2/3; (b) al final de la primaria; y (c) al final de la secundaria inferior.

Existe una amplia variedad de evaluaciones a nivel nacional, regional e internacional que producen información sobre el nivel de competencia en competencia lectora y matemáticas de niños y jóvenes. Estas evaluaciones internacionales a gran escala (ILSA por sus siglas en inglés) son consideradas por muchos como la fuente de información más adecuada para medir y monitorear el progreso hacia varios de los indicadores de la meta 4 de los ODS. Las ILSA ofrecen información sistemática de más de 100 sistemas educativos, brindan información sobre el nivel de competencia lectora y matemáticas de niños y jóvenes y son garantía de datos de calidad. Además, las ILSA producen datos que son estrictamente comparables entre países y a lo largo del tiempo. Sin embargo, cada una de estas evaluaciones tiene un marco de evaluación diferente, utilizan diferente escala y están diseñados para informar decisiones de política en diferentes contextos

Por lo tanto, aun cuando la información ya está disponible, para poder comparar los distintos programas de evaluación, es necesario que se integre en un marco global. Tal enfoque permitiría comparaciones entre sistemas educativos no solo dentro de una región geográfica sino globalmente y a lo largo del tiempo. permitiría su vez esto haría posible aprender de las iniciativas políticas de países con características similares (por ejemplo, en cuanto al nivel de desarrollo económico, estructura del sistema educativo) pero ubicados en diferentes partes del mundo. Si bien las similitudes entre los sistemas educativos son importantes para permitir comparaciones justas y válidas, las diferencias contextuales, históricas y culturales permiten la posibilidad de observar

1 Este resumen ejecutivo contiene extractos de los informes técnicos de Rosetta Stone proporcionados por el Centro de Estudios Internacionales TIMSS & PIRLS sobre el desarrollo de las tablas de concordancias TIMSS & PIRLS para ERCE y PASEC. Los informes están disponibles en <https://timssandpirls.bc.edu/Rosetta-Stone-Reports/index.html>, y otras preguntas deben dirigirse a timss@bc.edu.

2 Ver <https://en.unesco.org/education2030-sdg4/targets>

el mismo problema desde diversos ángulos y, por lo tanto, el desarrollo de enfoques y soluciones novedosos. Al mismo tiempo, las evaluaciones regionales y nacionales, al centrarse en una prueba creada para un contexto específico, facilitan el desarrollo de comparaciones más meticulosas que, a su vez, pueden informar las políticas de una manera más detallada y contextualizada a esa realidad regional y/o nacional. De esta manera, la combinación de evaluaciones nacionales, regionales y globales tiene el potencial de impulsar la profundidad y amplitud de nuestros esfuerzos para medir y monitorear el ODS 4.1.1.

Por estas razones, el Instituto de Estadística de la UNESCO (UIS por sus siglas en inglés) ha encabezado un programa metodológico que tiene como objetivo producir una estrategia de medición sólida que ofrezca información comparable para la mayor cantidad de países posible para medir y monitorear el progreso hacia el ODS 4.1.1. Este programa, pionero en su tipo, involucra dos evaluaciones regionales y dos globales, vinculadas de una manera que no requiere rediseñar en ningún aspecto las evaluaciones regionales, lo que proporciona la posibilidad de reajustar la información existente sin imponer decisiones desde arriba. Este es el espíritu que la UNESCO y otras agencias han tratado de mantener a lo largo del proceso del ODS: no obligar a los países a adoptar una evaluación única para reportar y medir progreso en la meta 1 del ODS 4.

Este programa, liderado por la Asociación Internacional para la Evaluación del Logro Educativo (IEA por sus siglas en inglés) y el Centro de Estudios Internacionales TIMSS & PIRLS, Lynch School of Education en Boston College, lleva el nombre del famoso descubrimiento arqueológico que permitió la traducción entre diferentes lenguajes escritos: Piedra de Rosetta (Rosetta Stone). El Estudio Piedra de Rosetta está diseñado para medir el progreso global hacia el ODS 4.1.1 al relacionar diferentes programas de evaluación regionales con el Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS por sus siglas en inglés) y el Estudio Internacional de Progreso en Competencia Lectora (PIRLS por sus siglas en inglés) métricas internacionales de larga data y puntos de referencia de logro³. El objetivo es proporcionar a los países que participaron en evaluaciones regionales o nacionales, pero no en TIMSS y PIRLS, información sobre las proporciones de estudiantes de primaria que han alcanzado un nivel mínimo de competencia en competencia lectora y matemáticas (ODS 4.1.1) que permita comparaciones internacionales bajo los mismos parámetros.

Este documento presenta un resumen ejecutivo de los primeros resultados del Estudio Piedra de Rosetta, consistente en el establecimiento de una tabla de concordancia que proyecta las distribuciones de puntaje estimadas a partir de dos evaluaciones regionales a las distribuciones en TIMSS y PIRLS. Estas evaluaciones son el Estudio Regional Comparativo y Explicativo de la UNESCO (ERCE) en los países de América Latina y el Caribe y el Programa para el Análisis de los Sistemas Educativos (PASEC; Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN) en países francófonos del África subsahariana.

El resto de este documento se divide en tres secciones principales. La primera sección describe los métodos y análisis realizados para establecer la concordancia entre las evaluaciones regionales y TIMSS y PIRLS. Las tablas de concordancia resultantes y las pautas para su uso e interpretación se presentan en la segunda sección. La tercera sección muestra la proporción de estudiantes que alcanzan el ODS 4.1.1 en cada uno de los países ERCE y PASEC según las tablas de concordancia elaboradas por la Piedra Rosetta. La última sección ofrece un conjunto de mensajes clave para formuladores de políticas, profesionales y académicos.

3 Véase IEA, 2017, en *Lecturas adicionales*.

MÉTODOS

A pesar de tener diferentes marcos y enfoques, las ILSA como ERCE, PASEC, TIMSS y PIRLS comparten un conjunto de procedimientos y métodos muy similares. Por ejemplo, las ILSA utilizan un muestreo complejo (diseño de muestreo estratificado, multietapas y agrupado) y una evaluación compleja (diseño de bloques rotados). Como resultado, el proceso de producir una concordancia entre ellos también es complejo. Esta sección comienza con una breve descripción de los procedimientos adoptados por las ILSA, seguida de una descripción general de los métodos utilizados para producir las tablas de concordancia. Es necesario hacer hincapié en que este documento ofrece una versión simplificada de ambos; se recomienda enfáticamente consultar los informes técnicos de las ILSA correspondientes y los informes de análisis del Estudio Piedra Rosetta para conocer los detalles técnicos completos (consulte la lista de lecturas sugeridas al final de este documento).

A. El diseño complejo de las ILSA

El objetivo de la mayoría de las ILSA es producir información representativa a nivel nacional sobre el logro educativo de una población estudiantil objetivo en una materia determinada. Para hacer esto, las ILSA enfrentan dos desafíos prácticos principales. En primer lugar, no se puede administrar una prueba a todos los estudiantes de un sistema educativo y, en segundo lugar, un solo estudiante no puede responder la prueba completa. Evaluar a todos los estudiantes sería muy costoso y crearía muchas complicaciones logísticas, y aquellos estudiantes que son evaluados no responden todos los ítems porque les llevaría entre ocho y diez horas completar una prueba que cubriera todos los temas incluidos en el marco de estudio. Por lo tanto, solo una muestra de estudiantes responde la prueba, y cada uno de los estudiantes de la muestra responde solo una parte de la prueba. Estos desafíos se superan mediante el uso de una muestra y un diseño de evaluación complejo.

El diseño de muestreo de las ILSA se caracteriza típicamente como un muestreo por conglomerados estratificado en varias etapas. La estratificación consiste en ordenar las listas de escuelas en grupos (es decir, estratos) por algunas de las características que comparten, como ubicación en regiones geográficas o tipos de escuelas (p. ej., públicas, privadas). La primera etapa de muestreo involucra la selección de escuelas dentro de estos estratos. La lista de escuelas también está ordenada según el tamaño de las escuelas, de manera que las escuelas más grandes (es decir, las que tienen más alumnos) tienen una mayor probabilidad de ser seleccionadas (probabilidad proporcional al tamaño). La segunda etapa involucra el muestreo de estudiantes o clases (es decir, grupos) dentro de las escuelas muestreadas. Dado que el objetivo es proporcionar estimaciones para toda la población objetivo, estas probabilidades desiguales deben ajustarse utilizando muestreo y pesos replicados (otros métodos, como los modelos multinivel, también se pueden usar en combinación con las ponderaciones de muestreo).

Las ILSA necesitan un diseño de evaluación complejo porque se requiere una gran cantidad de ítems (preguntas en el examen) para evaluar de manera confiable la competencia de los estudiantes en un dominio de

contenido tan amplio como puede ser matemáticas o lectura. Para incluir muchos ítems cuando se tiene un tiempo de prueba limitado, se utiliza un diseño de bloques rotados. En otras palabras, los ítems se agrupan en bloques de acuerdo al contenido y a los dominios cognitivos que miden. Luego, los bloques se distribuyen en varias pruebas, de modo que cada prueba contenga dos o más bloques, de modo que ninguna prueba contiene todos los ítems. La evaluación está diseñada de tal manera que las diferentes pruebas resultantes están vinculadas a través de bloques de ítems comunes. Como tal, ningún estudiante contesta todos los ítems y muchos estudiantes contestan ítems diferentes. Sin embargo, aún se puede estimar un puntaje de logro comparable gracias a un método de escalamiento conocido como teoría de respuesta al ítem (IRT por sus siglas en inglés). A través de la IRT, se estima la habilidad de cada persona y la dificultad de cada ítem. Esto permite estimar una puntuación que es comparable entre los estudiantes, incluso cuando respondieron diferentes ítems de la prueba. Luego, para estimar los puntajes de logro, se utiliza un modelo de población. Este modelo toma las estimaciones del modelo IRT y la información de contexto disponible de los estudiantes para estimar el puntaje que cada estudiante habría recibido si hubiera tomado todos los ítems de la prueba. Debido a que este puntaje se expresa como una distribución de probabilidad, se toman varios puntos aleatorios de la distribución (generalmente cinco). Cada punto aleatorio, también conocido como valor plausible (PV), se considera un valor representativo de la distribución de puntajes potenciales para todos los estudiantes en la muestra que tienen características de contexto similares y patrones similares de respuestas a los ítems (por ejemplo, aquellos que tienden a responder correctamente a elementos de la misma dificultad).

Tanto el muestreo como el diseño de evaluación dan como resultado cierta incertidumbre o error de medición que debe tenerse en cuenta en todos los cálculos. Es importante señalar que ningún diseño de evaluación elimina por completo el error de medición. En la práctica, se trata de optimizar la incertidumbre causada por el error de medición en relación con el costo y otras consideraciones prácticas. Esta incertidumbre se tiene en cuenta en todos los análisis mediante el uso de pesos de muestreo y pesos de replicación (para atender el diseño de muestra complejo) y mediante el uso de métodos IRT y valores plausibles (PV por sus siglas en inglés) (para atender el diseño de evaluación complejo).

B. El Método de la Piedra de Rosetta

Esta sección describe los instrumentos y el diseño del estudio de vinculación Piedra Rosetta y el procedimiento analítico seguido para la construcción de las tablas de concordancia. El estudio Piedra Rosetta, como la mayoría de las ILSA, utiliza un diseño de evaluación complejo en el que a cada estudiante se le administra solo un subconjunto de los ítems de la prueba. El Estudio Piedra Rosetta se compone de una evaluación en dos partes. La primera parte es la evaluación de los estudios fuente (es decir, ERCE y PASEC), incluyendo los ítems para medir logro y el cuestionario de contexto. La segunda parte es la pieza central del estudio, la evaluación Piedra Rosetta (estudio objetivo), que consta de pruebas con bloques de ítems de TIMSS y PIRLS. En total, se utilizaron ocho bloques rotados con ítems de matemáticas (TIMSS) y cuatro pasajes rotados con ítems de competencia lectora (PIRLS). Ambas partes de la evaluación se administraron en papel y a los mismos estudiantes. A cada estudiante se le administraron las pruebas del estudio de origen (PASEC o ERCE) el primer día y la prueba Piedra Rosetta el segundo día.

Para establecer tablas de concordancia, el análisis de los datos se realizó en cuatro pasos. Primero, se evaluó la calidad de los datos con base en estadísticas descriptivas y la variabilidad de la no respuesta. En segundo lugar, se usaron modelos IRT para construir escalas ERCE, PASEC y Piedra Rosetta (es decir, TIMSS y PIRLS) comparables entre las poblaciones de estudiantes. En tercer lugar, la metodología PV se utilizó por separado para ERCE, PASEC y Piedra Rosetta para producir puntajes de matemáticas y competencia lectora. En cuarto lugar, se establecieron tablas de concordancia basadas en los puntajes promedio en matemáticas y competencia lectora de los estudios fuente (ERCE y PASEC) y del estudio objetivo (Piedra Rosetta - TIMSS y PIRLS). El análisis se realizó con datos de tres países PASEC (Burundi, Guinea y Senegal) y de dos países ERCE (Colombia y Guatemala) usando los pesos muestrales que los equipos ERCE y PASEC proporcionaron al Centro de Estudios Internacionales TIMSS & PIRLS.

i. Evaluación de la calidad de los datos

La calidad de los datos se evaluó utilizando estadísticas clásicas sobre los ítems. La dificultad de los ítems se evaluó estimando el porcentaje de estudiantes que respondieron correctamente cada ítem, y la discriminación de ítems estimando la correlación biserial puntual entre el éxito en el ítem y la puntuación total. Los resultados de esta evaluación confirmaron que los datos recolectados a través del Estudio Piedra Rosetta tenían propiedades adecuadas para proceder a los siguientes pasos del análisis. La dificultad de los ítems estuvo generalmente dentro de un rango aceptable (25% a 95% correcto) y fue similar para los ítems ERCE, PASEC, TIMSS y PIRLS. Los valores de discriminación también fueron similares en todos los estudios y estuvieron dentro de rangos aceptables (coeficientes de correlación no inferiores a 0,20), con valores típicos entre 0,24 y 0,41 para PASEC y entre 0,36 y 0,50 para ERCE.

ii. Construcción de escalas comparables para ERCE, PASEC y Piedra Rosetta (es decir, TIMSS y PIRLS)

Se utilizaron modelos IRT unidimensionales para escalar los ítems de enlace de Piedra Rosetta (es decir, ítems TIMSS y PIRLS), así como los ítems PASEC y ERCE (por separado para matemáticas y comprensión lectora). Los modelos IRT unidimensionales mostraron altos niveles de comparabilidad (es decir, altas proporciones de ítems con buen ajuste) entre los países ERCE y las escalas Piedra Rosetta (con valores que oscilan entre 91,2 % y 95,1 %), así como entre los países PASEC y las escalas Piedra Rosetta (con valores que van del 81,2% al 86,3%). Los modelos también mostraron altos niveles de comparabilidad entre países para las escalas ERCE y PASEC (con valores que van desde 91,5% a 98,7%), proporcionando una base sólida para establecer una concordancia.

Después de escalar los ítems de PASEC, ERCE y Piedra Rosetta con modelos IRT unidimensionales separados, se utilizaron modelos IRT multidimensionales para examinar qué tan similares o diferentes son los constructos medidos de las diferentes evaluaciones. Más precisamente, estos modelos multidimensionales se utilizaron para investigar la relación entre las escalas ERCE y PASEC de matemáticas y TIMSS, y entre las escalas de competencia lectora de ERCE y PASEC y PIRLS. Los parámetros de los ítems en estos modelos multidimensionales se fijaron a los valores de los parámetros de los ítems obtenidos del conjunto de modelos unidimensionales descritos anteriormente.

Los resultados de los modelos IRT multidimensionales indicaron que las escalas correspondientes de la Piedra Rosetta y PASEC, y la Piedra Rosetta y ERCE miden constructos que no son idénticos (como se esperaba, ya que se desarrollaron en base a diferentes marcos) pero que son lo suficientemente similares (con correlaciones latentes entre dimensiones superiores a 0,80 e inferiores a 0,90) para permitir una concordancia significativa para una proyección de distribuciones de puntuación. La correlación latente osciló entre 0,82 y 0,90 para ERCE y entre 0,81 y 0,83 para PASEC.

iii. Producción de puntajes de matemáticas y comprensión lectora

Una vez que se estimaron los puntajes de IRT, se utilizó un modelo de población para estimar la distribución de competencias o PV. Este modelo de población utiliza las estimaciones del modelo IRT y la información de contexto disponible de los estudiantes, para estimar una distribución de los puntajes posibles que cada estudiante habría recibido si hubiera contestado la prueba completa. El objetivo final del modelado de población es producir distribuciones de competencias a partir de las cuales se pueden extraer PVs. Siguiendo los procedimientos utilizados en TIMSS y PIRLS, se extrajeron cinco PVs de la distribución de competencia para cada dominio y cada estudiante.

iv. Establecimiento de la concordancia

El paso final fue construir las tablas de concordancia de la Piedra Rosetta, que establecen una relación entre las puntuaciones de las evaluaciones de origen y de destino. En la tabla Piedra Rosetta, se proyecta un rango de puntajes proyectos de TIMSS y PIRLS a partir de puntajes de matemáticas y competencia lectora de ERCE y PASEC, respectivamente. Es decir, Matemáticas ERCE, competencia lectora ERCE, Matemáticas PASEC y competencia lectora PASEC representan las evaluaciones fuente y TIMSS y PIRLS representan las evaluaciones objetivo.

El primer paso fue evaluar la relación entre los datos de las evaluaciones de origen y de destino. Los análisis de correlación entre los puntajes promedio de matemáticas ERCE y TIMSS ($r = 0,80$ a $0,82$), matemáticas PASEC y TIMSS ($r = 0,70$ a $0,80$), competencia lectora ERCE y PIRLS ($r = 0,78$ a $0,82$) y competencia lectora PASEC y PIRLS ($r = 0,73$ a $0,81$) indicaron que las escalas ERCE y la Piedra Rosetta y las escalas PASEC y la Piedra Rosetta miden constructos diferentes pero similares; es decir, las correlaciones fueron razonablemente altas para construir una concordancia.

Los puntajes y niveles de concordancia se identificaron luego con base en los puntajes promedio estimados de PASEC y ERCE usando los datos combinados de Burundi, Guinea y Senegal para PASEC y Colombia y Guatemala para ERCE. Los rangos de puntuación de las puntuaciones medias de las escalas de matemáticas y competencia lectora PASEC y ERCE se redondearon hacia arriba o hacia abajo para cubrir casi todos los datos de los países participantes, y para ser lo más simétricos posible alrededor de la media general (500 para PASEC, 700 para ERCE). Para ambas escalas PASEC, matemáticas y competencia lectora, los puntajes oscilaron entre 200 y 800. Para ambas escalas ERCE, matemáticas y competencia lectora, los puntajes oscilaron entre 440 y 940. En ambos casos, estos rangos cubren casi el 99,5% de los datos.

Tanto para las escalas ERCE como PASEC, se especificaron 20 puntos como el intervalo de puntaje para incluir suficientes puntajes o niveles de competencia y retener la mayor cantidad de información posible. Como resultado, hay 26 niveles de puntuación dentro del rango de puntuación de 440 a 940 y 200 y 800, respectivamente.

Para cada nivel de puntaje de concordancia identificado, se utilizó un modelo predictivo de emparejamiento de medias (PMM por sus siglas en inglés) para seleccionar cinco donantes (es decir, estudiantes examinados) para cada prueba y de cada país. De modo que cada país contribuyera por igual a cada una de las tablas de concordancia. Cada uno de los donantes donó cinco PVs en las pruebas objetivo. La media y la desviación estándar de los PVs de los donantes a partir de los datos de vinculación de Piedra Rosetta se calcularon en función del total de PVs donados en cada nivel de concordancia. Téngase en cuenta que estos pasos se implementaron por separado para matemáticas y competencia lectora de ERCE y para matemáticas y competencia lectora de PASEC.

Se crearon tablas de concordancia preliminares para matemáticas ERCE, competencia lectora ERCE, matemáticas PASEC y competencia lectora PASEC mediante la asignación de la media estimada y la desviación estándar de cada conjunto de PVs en función de los datos de vinculación de la Piedra Rosetta (TIMSS y PIRLS), respectivamente, a cada nivel de puntuación de concordancia en el rango especificado de matemáticas ERCE, competencia lectora ERCE, matemáticas PASEC y competencia lectora PASEC. Finalmente, para cada punto de puntuación de concordancia, la media de los PVs donados se suavizó aplicando un promedio móvil simple, utilizando una ventana de siete puntos de puntuación. La desviación estándar de los PVs de cada punto de puntuación se suavizó de manera similar a las medias de los PVs, usando una media geométrica móvil de las varianzas de cada conjunto de las cinco medias de PVs donadas agrupadas en el nivel de puntuación correspondiente en la tabla. La raíz cuadrada de esta varianza suavizada se convierte en la desviación estándar condicional suavizada.

LASTABLAS DE CONCORDANCIA

Las tablas 1 y 2 muestran las tablas de concordancia final para matemáticas ERCE/PASEC y comprensión lectora ERCE/PASEC, respectivamente. La primera y quinta columna de cada tabla muestran los niveles de puntuación de concordancia de la evaluación de origen (ERCE o PASEC), ya sea en matemáticas o comprensión lectora. Las columnas segunda y sexta muestran la distribución predictiva, las medias proyectadas y las desviaciones estándar condicionales de la puntuación de la evaluación de la fuente en la escala TIMSS o PIRLS. Las columnas tercera, cuarta, séptima y octava muestran los valores de escala para los percentiles de esta distribución condicional dada la evaluación de la fuente (ERCE o PASEC) en las escalas TIMSS o PIRLS. Se proporcionan los valores inferior y superior en los percentiles 68 y 95.

TABLA 1. Tabla de concordancia para matemáticas ERCE y PASEC

Puntaje de Matemáticas ERCE	Puntaje Proyectado en la Escala TIMSS		Límite inferior		Límite superior	
	Media	DE	95%	68%	68%	95%
400	290	64	162	226	354	417
420	304	63	178	241	367	430
440	319	62	194	256	381	443
460	318	63	192	255	381	444
480	322	62	198	260	384	446
500	326	62	201	264	389	451
520	334	62	211	273	396	458
540	342	64	214	278	406	470
560	357	63	231	294	419	482
580	371	62	247	309	433	495
600	389	61	266	327	450	511
620	403	61	282	342	463	524
640	420	58	303	361	478	537
660	432	57	317	375	489	546
680	449	53	344	397	502	555
700	465	52	362	414	517	569
720	481	51	379	430	532	583
740	497	49	399	448	547	596
760	515	50	415	465	565	616
780	531	50	431	481	581	631
800	548	48	453	500	596	643
820	563	46	471	517	609	655
840	576	46	484	530	622	668
860	590	45	500	545	635	680
880	599	46	508	554	645	691
900	608	46	516	562	654	699
920	617	48	520	568	665	713
940	624	51	522	573	675	726
960	638	50	538	588	688	739
980	653	49	554	603	702	751

Puntaje de Matemáticas ERCE	Puntaje Proyectado en la Escala TIMSS		Límite inferior		Límite superior	
	Media	DE	95%	68%	68%	95%
220	197	76	44	121	274	350
240	210	76	58	134	286	362
260	223	75	72	148	298	374
280	224	75	74	149	300	375
300	229	77	75	152	306	382
320	231	76	78	155	307	383
340	235	75	85	160	310	385
360	241	71	99	170	312	383
380	254	69	115	184	323	392
400	265	68	130	197	333	401
420	273	68	137	205	341	410
440	284	65	154	219	349	414
460	297	62	172	234	359	421
480	315	61	194	254	376	437
500	336	63	209	273	399	462
520	344	65	215	280	409	473
540	355	64	227	291	419	483
560	371	63	245	308	434	497
580	382	66	251	317	448	514
600	395	70	256	326	465	535
620	403	71	260	332	475	546
640	417	71	274	345	488	559
660	437	69	299	368	506	575
680	453	69	316	385	522	591
700	469	67	335	402	536	602
720	484	64	357	421	548	612
740	500	57	386	443	556	613
760	513	52	408	461	566	618
780	526	52	422	474	578	630
800	539	51	436	487	590	641

DE = desviación estándar

TABLA 2. Tabla de concordancia para comprensión lectora ERCE y PASEC

Puntaje de comprensión lectora ERCE	Puntaje Proyectado en la Esca-la PIRLS		Límite inferior		Límite superior	
	Media	DE	95%	68%	68%	95%
400	284	84	116	200	369	453
420	296	83	130	213	380	463
440	309	82	144	226	391	473
460	321	81	158	240	402	483
480	329	80	169	249	409	489
500	342	77	187	265	419	497
520	347	76	195	271	424	500
540	373	74	224	298	447	521
560	378	74	229	303	452	526
580	395	68	258	326	463	532
600	410	65	279	344	475	540
620	421	62	297	359	483	545
640	436	62	311	373	498	560
660	455	63	328	392	518	581
680	467	64	338	402	531	595
700	482	60	362	422	542	602
720	488	63	361	425	551	615
740	501	65	372	437	566	631
760	522	64	393	457	586	650
780	526	63	399	463	590	653
800	536	61	414	475	597	658
820	546	60	425	486	607	667
840	555	60	436	495	614	674
860	565	59	447	506	623	682
880	568	56	456	512	624	680
900	576	56	463	519	632	688
920	588	58	473	530	646	703
940	596	59	477	537	655	715
960	608	58	492	550	666	725
980	620	57	506	563	678	735

Puntaje de comprensión lectora ERCE	Puntaje Proyectado en la Esca-la PIRLS		Límite inferior		Límite superior	
	Media	DE	95%	68%	68%	95%
220	146	72	2	74	218	290
240	161	72	17	89	233	304
260	175	72	31	103	247	319
280	178	72	34	106	249	321
300	181	72	38	110	253	325
320	190	71	47	118	261	332
340	196	72	52	124	267	339
360	205	71	63	134	276	347
380	216	72	73	145	288	359
400	228	72	84	156	300	372
420	238	76	87	163	314	390
440	253	74	104	179	327	401
460	265	73	120	193	338	411
480	280	71	139	209	351	422
500	297	71	155	226	369	440
520	317	73	172	244	390	462
540	330	72	186	258	402	474
560	351	66	219	285	417	482
580	364	66	232	298	430	496
600	377	68	241	309	446	514
620	392	69	255	323	461	529
640	405	67	271	338	471	538
660	420	63	295	357	483	545
680	444	66	312	378	511	577
700	456	69	319	388	525	593
720	473	71	332	402	544	615
740	486	70	346	416	555	625
760	492	72	347	420	565	637
780	507	72	362	434	579	651
800	521	72	377	449	593	665

DE = desviación estándar

Las tablas de concordancia deben interpretarse con cautela, ya que no son predicciones perfectas de cómo se desempeñaría un estudiante en una prueba específica (p. ej., TIMSS o PIRLS). No proporcionan un vínculo directo entre las evaluaciones y dependen de las características de la muestra. Por lo tanto, la incertidumbre de la predicción debe tenerse en cuenta al utilizar e interpretar las tablas de concordancia. Por ejemplo, un puntaje de matemáticas PASEC de 500 no resulta en un puntaje TIMSS de 336 sino en una distribución de puntajes TIMSS con una media de 336. En otras palabras, suponiendo distribuciones de puntuación condicional aproximadamente normales, el 68 % de los PVs generados en la escala TIMSS probablemente caerían en el rango de puntuación de 273 a 399 (si un alumno con una capacidad similar realizara la evaluación TIMSS) y el 95 % de los PVs generados en la escala TIMSS probablemente caería en el rango de puntuación de 209 a 462, como se muestra en la Tabla 1.

PROPORCIÓN DE ESTUDIANTES QUE LOGRAN EL ODS 4.1.1

Además de hacer inferencias sobre el rango de puntaje probable en las escalas TIMSS o PIRLS dado un puntaje ERCE o PASEC, también es posible generar los PVs probables para estudiantes individuales (que participaron en ERCE o PASEC pero no en TIMSS o PIRLS) en TIMSS y PIRLS escala utilizando las medias proyectadas y las desviaciones estándar de las tablas de concordancia. Los informes de análisis de la Piedra Rosetta brindan instrucciones detalladas sobre cómo generar PVs aleatorios utilizando las tablas de concordancia. Estos PVs proyectados luego se pueden usar para estimar el porcentaje de estudiantes que alcanzan un punto de referencia determinado, por ejemplo, el nivel mínimo de competencia global (MPL por sus siglas en inglés) para el final de la primaria, que indica el logro del ODS 4.1.1.

Cabe señalar que los niveles de logro de ERCE y PASEC que sirven como MPLs están destinados a mapear en el MPL global y, por lo tanto, a indicar el logro de ODS 4.1.1. Sin embargo, fueron fijados de manera independiente por los equipos técnicos de cada evaluación. En consecuencia, son cualitativamente diferentes entre sí. Las descripciones de lo que los estudiantes saben y pueden hacer cuando alcanzan los MPLs establecidos para ERCE y PASEC, así como para el MPL global (establecido por consenso de expertos internacionales y alineado con los puntos de referencia de TIMSS/PIRLS)⁴ se pueden encontrar en la Tabla 3.

4 Consulte ACER, 2019 y UIS, 2022 en *Lecturas adicionales*.

TABLA 3. Definición de los puntos de referencia ERCE, PASEC y TIMSS/PIRLS que sirven como MPLs para Grado 6 (ERCE y PASEC) y Grado 4 (TIMSS/PIRLS)

Estudio	Matemáticas		Comprensión lectora	
ERCE	Nivel 3 (789 puntos)	Los estudiantes de este nivel mostraron evidencia de ser capaces de: • Resolver problemas más complejos que requieran interpretar información e involucren dos o más operaciones incluyendo multiplicación o división. • Interpretar el significado de variaciones proporcionales en situaciones contextualizadas. • Identificar fracciones equivalentes (con denominador distinto de 10) y calcular sumas y restas de fracciones con el mismo denominador. • Relacionar números decimales con fracciones propias simples o números mixtos simples (p. ej., con denominador 2) y calcular o estimar sumas y restas de números decimales. • Determinar los términos intermedios faltantes de una sucesión presentada en una situación contextualizada, interpretando su patrón de formación. • Identificar relaciones de perpendicularidad y paralelismo en el plano. • Resolver problemas complejos que impliquen cálculo o estimación de áreas y perímetros de figuras geométricas. • Resolver problemas de medidas (masa, volumen y tiempo) y convertir unidades de medida. • Resolver problemas que requieran leer e interpretar información de tablas y gráficos o identificar gráficos que representen información entregada en diferentes formatos. Los estudiantes de este nivel dominan una secuencia numérica oral (contar hasta 60 en dos minutos) y son capaces de comparar números, completar series lógicas y realizar operaciones (sumas y restas) con números mayores de 50. Pueden resolver problemas con números menores de 20 utilizando habilidades de razonamiento.	Nivel 3 (754 puntos)	Los estudiantes de este nivel mostraron evidencia de ser capaces de: • Hacer inferencias a partir de conexiones entre ideas específicas o secundarias y ubicadas en diferentes partes de uno o más textos. • Inferir el tema central de un párrafo o parte del texto, estableciendo una relación con el texto en su conjunto. • Hacer inferencias (p. ej., inferir el tema central, las características o sentimientos de los personajes, el conflicto y el desenlace) que requieran comprender el texto de forma global e integrar las ideas implícitas presentes en él. • Interpretar expresiones en lenguaje figurado a partir de claves que están implícitas en el texto o que desafían el conocimiento del alumno sobre el mundo que le rodea. • Relacionar información visual y verbal en un texto.
PASEC	Nivel 3 (609 puntos)	Los estudiantes de este nivel dominan una secuencia numérica oral (contar hasta 60 en dos minutos) y son capaces de comparar números, completar series lógicas y realizar operaciones (sumas y restas) con números mayores de 50. Pueden resolver problemas con números menores de 20 utilizando habilidades de razonamiento.	Nivel 4 (595 puntos)	Lector intermedio: la mayor autonomía de comprensión lectora está reforzando la competencia de oraciones y textos por parte de los estudiantes. Los estudiantes han adquirido competencias de decodificación del lenguaje escrito y competencia auditiva que les permiten comprender información explícita en palabras, oraciones y pasajes breves. Pueden combinar sus habilidades de decodificación y su dominio del lenguaje oral para comprender el significado literal de un pasaje breve.
TIMSS/PIRLS	TIMSS Inter-medio (475 puntos)	Los estudiantes pueden aplicar conocimientos matemáticos básicos en situaciones sencillas. Pueden calcular con números enteros de tres y cuatro dígitos en una variedad de situaciones. Tienen cierta competencia de decimales y fracciones. Los estudiantes pueden identificar y dibujar formas con propiedades simples. Pueden leer, etiquetar e interpretar información en gráficos y tablas.	PIRLS Bajo (400 puntos)	Al leer textos literarios predominantemente más simples, los estudiantes pueden: • Localizar y recuperar información, acciones o ideas declaradas explícitamente. • Hacer inferencias directas sobre eventos y razones para las acciones. • Comenzar a interpretar los eventos de la historia y las ideas centrales. Al leer textos informativos predominantemente más simples, los estudiantes pueden: • Localizar y reproducir información declarada explícitamente de texto y otros formatos (p. ej., gráficos, diagramas). • Comenzar a hacer inferencias directas sobre explicaciones, acciones y descripciones.

Las proporciones de estudiantes en los países ERCE y PASEC que alcanzan el ODS 4.1.1 para Matemáticas (punto de referencia intermedio en TIMSS) según ERCE (Nivel de rendimiento 3) y PASEC (Nivel de rendimiento 3), así como según las tablas de concordancia producidas por el Estudio Rosetta Stone, se muestran en la Tabla 4. Las proporciones de estudiantes en los países ERCE y PASEC que alcanzan el ODS 4.1.1 para competencia lectora (punto de referencia bajo en PIRLS) según ERCE (Nivel de desempeño 3) y PASEC (Nivel de desempeño 4), así como de acuerdo con las tablas de concordancia producidas por el Estudio Piedra Rosetta, se muestran en la Tabla 5. Los errores estándar de estos porcentajes se calculan utilizando pesos adecuados y fórmulas de PV para incluir errores de evaluación y muestreo (consulte el informe de análisis de la Piedra Rosetta para obtener más detalles).

Es importante señalar que los porcentajes estimados con base al estudio Piedra Rosetta son en muchos casos considerablemente diferentes a los porcentajes reportados con base en los puntajes de PASEC y ERCE. En la mayoría de los casos, los porcentajes son mayores cuando las estimaciones se basan en Rosetta Stone para ERCE y menores para PASEC. Estas discrepancias podrían deberse a diferencias en los marcos de evaluación y los constructos medidos en cada uno. Como se mencionó anteriormente, los MPLs establecidos por cada evaluación son sustancialmente diferentes. Por ejemplo, mientras que ERCE considera que se ha alcanzado el MPL cuando los estudiantes pueden “interpretar expresiones en lenguaje figurado a partir de pistas que están implícitas en el texto”, PASEC considera que se ha alcanzado el MPL cuando los estudiantes pueden “[...] combinar sus habilidades de decodificación y su dominio del lenguaje oral para captar el significado literal de un pasaje corto” (ver Tabla 3).

Además, PASEC se desarrolló dentro y para el contexto de los países francófonos del África subsahariana, mientras que ERCE se desarrolló dentro y para el contexto de los países latinoamericanos. Por otro lado, TIMSS y PIRLS fueron diseñados para ser iniciativas globales. Si bien estas evaluaciones miden habilidades en matemáticas y competencia lectora, la definición misma de los constructos varía entre ellas; tienen diferentes puntos de partida y establecen diferentes aspiraciones para sus sistemas educativos.

Finalmente, las poblaciones objetivo de las evaluaciones son diferentes: Grado 6 para PASEC y ERCE y Grado 4 para TIMSS y PIRLS. Otro factor que también podría jugar un papel importante en la explicación de estas diferencias, es la etapa en la que se introduce la transición de los idiomas locales al idioma de prueba (por ejemplo, francés, español) en el sistema escolar. En muchos países, esta transición ocurre en el primer grado, mientras que en otros es tan tarde como el cuarto grado. En este punto, todas estas son hipótesis no verificadas y se necesita más investigación (y muestras más grandes de países y estudiantes que participen en Rosetta Stone) para ponerlas a prueba y dar respuestas claras.

TABLA 4. Proporción de estudiantes en países ERCE y PASEC que logran el ODS 4.1.1 en Matemáticas

País	Estudiantes por encima del MPL General		Estudiantes por encima del MPL Niños		Estudiantes por encima del MPL Niñas	
	%	EE	%	EE	%	EE
MPL de ERCE (6º grado)						
Argentina	13.1	(0.9)	14.2	(1.1)	12.0	(0.9)
Brasil	28.7	(1.5)	30.5	(1.7)	26.7	(1.7)
Colombia	16.5	(1.2)	17.6	(1.4)	15.3	(1.7)
Costa Rica	20.8	(1.7)	22.6	(1.9)	19.0	(1.9)
Cuba	20.6	(1.5)	20.5	(1.6)	20.8	(1.7)
Rep. Dominicana	2.1	(0.5)	2.1	(0.5)	2.1	(0.6)
Ecuador	22.8	(1.4)	23.0	(1.5)	22.4	(1.6)
El Salvador	7.4	(0.6)	8.0	(0.7)	6.8	(0.8)
Guatemala	6.6	(0.9)	7.6	(1.2)	5.6	(0.8)
Honduras	11.1	(1.9)	11.9	(2.0)	10.4	(1.9)
México	37.9	(1.3)	36.9	(1.6)	38.9	(1.7)
Nicaragua	3.0	(0.5)	3.8	(0.7)	2.3	(0.5)
Panamá	3.3	(0.4)	3.4	(0.5)	3.2	(0.5)
Paraguay	5.6	(0.7)	5.8	(0.8)	5.3	(0.9)
Perú	38.7	(1.5)	38.3	(1.8)	39.2	(1.8)
Uruguay	37.8	(1.5)	38.1	(1.8)	37.5	(1.7)
MPL de PASEC (6º grado)						
Benín	19.3	(2.6)	18.9	(2.4)	19.8	(3.0)
Burkina Faso	25.2	(1.2)	26.5	(1.5)	24.1	(1.4)
Burundi	18.2	(1.5)	23.7	(1.8)	13.8	(1.5)
Camerún	11.2	(1.0)	10.5	(1.0)	12.1	(1.4)
Chad	1.9	(0.6)	1.8	(0.6)	2.0	(0.7)
Congo	7.8	(0.9)	6.9	(1.0)	8.6	(1.3)
Costa de Marfil	2.6	(0.6)	2.9	(0.7)	2.3	(0.7)
Gabón	23.1	(1.9)	26.5	(2.4)	19.9	(2.0)
Guinea	6.9	(1.2)	7.5	(1.4)	6.2	(1.4)
Madagascar	6.3	(1.7)	5.7	(1.7)	7.0	(1.9)
Níger	8.0	(1.3)	8.1	(1.5)	7.8	(1.4)
R. D. del Congo	3.2	(1.1)	3.4	(1.3)	3.0	(1.0)
Senegal	27.6	(1.9)	27.8	(2.2)	27.4	(2.2)
Togo	16.0	(1.2)	16.1	(1.4)	15.8	(1.3)

	Estudiantes por encima del MPL General		Estudiantes por encima del MPL Niños		Estudiantes por encima del MPL Niñas	
	%	EE	%	EE	%	EE
MPL de Rosetta Stone (PIRLS 4º grado)						
	42.7	(1.4)	44.2	(1.6)	41.2	(1.6)
	56.9	(1.5)	58.2	(1.9)	55.6	(1.8)
	48.2	(1.8)	49.4	(2.1)	46.9	(2.1)
	56.4	(1.6)	57.9	(2.2)	54.9	(1.9)
	42.2	(1.9)	42.0	(2.0)	42.7	(2.1)
	22.2	(1.4)	20.9	(1.4)	23.5	(2.0)
	53.1	(1.5)	53.1	(1.7)	53.0	(1.7)
	35.7	(1.0)	36.5	(1.7)	34.9	(1.6)
	30.1	(1.2)	32.4	(1.5)	27.8	(1.5)
	38.5	(2.2)	40.4	(2.3)	36.7	(2.6)
	66.1	(1.5)	64.9	(1.7)	67.3	(1.6)
	30.2	(1.2)	33.0	(1.8)	27.5	(1.5)
	25.2	(1.0)	25.4	(1.3)	25.0	(1.2)
	27.1	(1.3)	27.8	(1.6)	26.4	(1.6)
	66.3	(1.3)	66.1	(1.6)	66.6	(1.7)
	65.8	(1.4)	66.6	(2.0)	65.0	(1.9)
MPL de Rosetta Stone (PIRLS 4º grado)						
	8.9	(1.4)	8.8	(1.5)	9.0	(1.7)
	10.2	(0.8)	10.9	(0.9)	9.5	(1.1)
	7.8	(0.7)	9.6	(1.0)	6.4	(0.8)
	5.1	(0.8)	5.0	(0.9)	5.1	(1.0)
	1.2	(0.3)	1.3	(0.4)	1.0	(0.5)
	3.7	(0.5)	3.5	(0.7)	3.9	(0.7)
	1.5	(0.4)	1.6	(0.4)	1.4	(0.5)
	8.8	(1.1)	9.8	(1.8)	7.9	(1.1)
	3.1	(0.7)	3.6	(0.9)	2.5	(0.6)
	3.0	(0.7)	2.9	(0.8)	3.0	(0.8)
	3.5	(0.7)	3.6	(0.8)	3.4	(0.8)
	1.9	(0.7)	2.1	(0.9)	1.6	(0.6)
	12.0	(1.3)	12.6	(1.5)	11.5	(1.3)
	7.0	(0.6)	7.0	(0.7)	6.9	(0.7)

MPL = nivel mínimo de competencia por sus siglas en inglés; EE = error estándar

TABLE 5. Proporción de estudiantes en países ERCE y PASEC que logran el ODS 4.1.1 en compresion lectora

País	Estudiantes por encima del MPL General		Estudiantes por encima del MPL Niños		Estudiantes por encima del MPL Niñas	
	%	EE	%	EE	%	EE
MPL de ERCE (6° grado)						
Argentina	31.7	(1.2)	28.8	(1.3)	34.5	(1.4)
Brasil	43.3	(1.5)	40.2	(1.7)	46.7	(1.8)
Colombia	37.3	(1.7)	35.2	(2.2)	39.5	(2.0)
Costa Rica	53.8	(1.7)	51.1	(2.0)	56.4	(1.9)
Cuba	44.4	(1.3)	38.6	(1.6)	50.4	(1.6)
Rep. Dominicana	16.3	(1.1)	12.1	(1.0)	20.6	(1.5)
Ecuador	26.0	(1.0)	24.1	(1.2)	27.9	(1.3)
El Salvador	29.2	(1.2)	26.3	(1.3)	32.3	(1.6)
Guatemala	15.9	(1.0)	15.5	(1.4)	16.3	(1.1)
Honduras	16.1	(1.5)	14.0	(1.5)	18.1	(1.9)
México	41.4	(1.4)	37.8	(1.8)	45.0	(1.6)
Nicaragua	12.8	(0.8)	11.7	(0.8)	14.0	(1.1)
Panamá	17.4	(1.1)	15.3	(1.1)	19.4	(1.2)
Paraguay	18.7	(1.1)	15.8	(1.2)	21.5	(1.4)
Perú	48.8	(1.2)	44.6	(1.5)	53.4	(1.5)
Uruguay	43.5	(1.3)	40.3	(1.5)	46.7	(1.5)
MPL de PASEC (6° grado)						
Benín	45.5	(2.7)	44.0	(2.6)	47.3	(3.1)
Burkina Faso	33.0	(1.4)	32.6	(1.5)	33.3	(1.6)
Burundi	4.5	(1.0)	4.7	(1.0)	4.4	(1.1)
Camerún	30.3	(1.7)	28.2	(1.7)	32.7	(2.2)
Chad	7.6	(1.2)	7.8	(1.2)	7.4	(1.4)
Congo	33.6	(1.8)	30.2	(1.9)	37.1	(2.2)
Costa de Marfil	22.1	(1.9)	21.0	(2.2)	23.3	(1.9)
Gabón	76.3	(1.8)	73.7	(2.1)	78.9	(1.9)
Guinea	22.2	(1.7)	21.9	(1.9)	22.6	(2.1)
Madagascar	6.3	(2.0)	5.4	(1.8)	7.1	(2.4)
Níger	14.4	(1.6)	13.1	(1.8)	15.9	(1.7)
R. D. del Congo	9.2	(1.5)	9.9	(1.4)	8.5	(1.8)
Senegal	41.1	(2.2)	38.2	(2.3)	43.5	(2.7)
Togo	19.4	(1.1)	18.5	(1.3)	20.2	(1.3)

	Estudiantes por encima del MPL General		Estudiantes por encima del MPL Niños		Estudiantes por encima del MPL Niñas	
	%	EE	%	EE	%	EE
MPL de Rosetta Stone (PIRLS 4° grado)						
	78.9	(1.1)	77.6	(1.3)	80.2	(1.3)
	86.2	(0.8)	84.4	(1.1)	88.1	(1.0)
	83.7	(1.1)	82.3	(1.5)	85.1	(1.5)
	90.6	(0.9)	89.8	(0.9)	91.5	(1.3)
	86.6	(1.1)	83.6	(1.6)	89.6	(1.1)
	65.7	(1.2)	61.1	(1.7)	70.4	(1.6)
	76.5	(1.1)	75.9	(1.5)	77.1	(1.2)
	80.2	(1.1)	78.3	(1.3)	82.1	(1.3)
	67.0	(1.4)	67.0	(2.0)	66.9	(2.3)
	72.1	(1.5)	71.7	(1.8)	72.5	(1.6)
	84.1	(0.9)	83.0	(1.1)	85.1	(1.1)
	71.4	(1.2)	70.4	(1.7)	72.5	(1.5)
	68.7	(1.4)	67.2	(1.7)	70.0	(1.5)
	69.7	(1.3)	66.9	(1.6)	72.4	(1.6)
	85.7	(1.0)	84.5	(1.4)	87.1	(1.4)
	85.3	(0.9)	83.7	(1.0)	87.0	(1.3)
MPL de Rosetta Stone (PIRLS 4° grado)						
	37.4	(2.2)	36.2	(2.5)	38.7	(2.8)
	27.8	(1.0)	27.8	(1.5)	27.7	(1.5)
	9.9	(0.8)	10.5	(0.9)	9.4	(1.0)
	25.0	(1.3)	24.0	(1.5)	26.1	(1.8)
	8.8	(1.1)	9.2	(1.3)	8.2	(1.5)
	28.0	(1.5)	25.6	(1.9)	30.5	(1.9)
	19.3	(1.4)	19.2	(1.7)	19.5	(1.6)
	56.7	(1.8)	55.2	(2.3)	58.3	(2.2)
	19.5	(1.6)	19.3	(1.9)	19.7	(1.9)
	8.1	(1.2)	7.1	(1.3)	9.1	(1.4)
	13.4	(1.2)	12.6	(1.5)	14.4	(1.5)
	11.0	(1.2)	11.9	(1.4)	10.0	(1.4)
	34.2	(1.9)	32.9	(2.3)	35.3	(2.3)
	17.5	(0.9)	16.7	(1.2)	18.2	(1.3)

MPL = nivel mínimo de competencia por sus siglas en inglés; EE = error estándar

MENSAJES CLAVE

- El marco de los ODS exige explícitamente que los sistemas nacionales realicen un seguimiento del logro de sus objetivos utilizando indicadores comparables, es decir, indicadores que se puedan comparar entre países y a lo largo del tiempo. Sin embargo, los organismos de las Naciones Unidas se han mostrado reacios a imponer un marco de medición específico de arriba hacia abajo. Aparte de las cuestiones humanísticas y políticas involucradas, tal tarea requeriría que todos los involucrados invirtieran tiempo y recursos excesivos. Además, muchas teorías del desarrollo económico y las finanzas públicas sugieren minimizar las soluciones que son demasiado estandarizadas. Por ello se consideró que un compromiso óptimo era desarrollar escalas para las herramientas de medición existentes y hacerlas comparables ex-post, a través de medios analíticos y empíricos. Esto es lo que ha hecho el estudio Piedra Rosetta sosteniendo la perspectiva de que los contextos nacionales son diferentes (en el África subsahariana y en Latinoamérica) y que las comparaciones globales poseen sus limitaciones. Se trata de un esfuerzo pionero, quizás el primero de su tipo en el campo de la medición del aprendizaje. Esfuerzos similares, pero con diferencias importantes, se están llevando a cabo por UIS y otros socios.
- Se establecieron escalas comparables entre países y entre evaluaciones. Los resultados del estudio Piedra Rosetta indican que, desde una perspectiva estadística, los constructos de competencia lectora y matemáticas medidos por PASEC y ERCE son lo suficientemente similares a los constructos medidos por TIMSS y PIRLS para permitir la concordancia deseada. Las tablas de concordancia constituyen el primer paso hacia el entendimiento de las diferencias y deben ser parte de un ejercicio amplio en términos de países y muestras de escuelas para poder realizar inferencias robustas. Las tablas de concordancia incluidas en este documento se pueden utilizar para hacer inferencias sobre el rango de puntaje probable en las escalas TIMSS o PIRLS dado un puntaje en PASEC o ERCE, lo que permite a los países comparar el rendimiento de sus estudiantes a nivel mundial y proporciona datos que los ayudan a medir el progreso global hacia ODS 4.1.1.
- Los resultados del estudio Piedra Rosetta para ERCE y PASEC sugieren que se pueden establecer tablas de concordancia para otras evaluaciones regionales (p. ej., SAQMEC, SEA-PLM, PILNA). La información resultante sumado a más investigación y análisis sería extremadamente útil y valiosa, ya que permitiría comparar todas las evaluaciones regionales no solo con TIMSS y PIRLS, sino también entre sí.
- La información proporcionada por el estudio Piedra Rosetta y las evaluaciones regionales como PASEC y ERCE deben considerarse complementarias. Tanto las evaluaciones regionales como Piedra Rosetta juegan un papel insustituible en la estrategia global para medir y monitorear el progreso hacia el ODS 4.1.1. Juntos mejoran las posibilidades de profundidad y amplitud del análisis que se puede llevar a cabo y, en consecuencia, mejoran la calidad y relevancia de la información disponible para los formuladores de políticas y otras partes interesadas.

- El estudio Piedra Rosetta abre un sinfín de posibilidades para los análisis secundarios que pueden ayudar a mejorar los informes globales y facilitar los análisis comparativos de los sistemas educativos de todo el mundo. Cada evaluación refleja el currículo de la región/país, mientras que la distribución del nivel de competencia es adecuada para la presentación de informes globales.
- Las tablas de concordancia presentadas aquí se basan en solo dos países para ERCE y tres países para PASEC. Para los países que no participaron en el estudio Piedra Rosetta (es decir, no administraron las pruebas de enlace de TIMSS/PIRLS), el uso de las tablas de concordancia proporcionadas en este documento es una extrapolación y conlleva una incertidumbre adicional que no se puede magnificar sin realizar también una recopilación de datos con los instrumentos del estudio Piedra Rosetta. Sin embargo, y este punto no se puede enfatizar demasiado, este enfoque es mejor que cualquier otra alternativa existente.
- Aumentar el tamaño de las muestras nacionales y agregar más países por evaluación regional mejoraría aún más la concordancia estimada y permitiría realizar investigaciones para explicar las diferencias observadas en el porcentaje de estudiantes que alcanzan el ODS 4.1.1 cuando se estima con Piedra Rosetta versus ERCE o PASEC.
- Se necesita más reflexión sobre el establecimiento de los niveles mínimos de competencia para cada estudio regional que mejor se correspondan con el nivel de competencia global acordado para garantizar comparaciones más justas de los porcentajes de estudiantes que alcanzan el ODS 4.1.1 en cada sistema educativo.
- Si los países estuvieran interesados en establecer una concordancia entre su evaluación nacional y TIMSS y PIRLS, se podría aplicar el enfoque de Piedra Rosetta. Sin embargo, los costos por país serían relativamente altos ya que los resultados solo serían aplicables a un país y los instrumentos tendrían que adaptarse a cada evaluación nacional.
- Si bien agregar evaluaciones nacionales al estudio Piedra Rosetta podría ser una opción para los países que aún no están incluidos en ninguna evaluación regional, debido a la complementariedad de las evaluaciones regionales y con el estudio Piedra Rosetta discutida anteriormente, está claro que tener más países participando en ambas evaluaciones regionales (por ejemplo, PASEC y ERCE) e internacionales (por ejemplo, TIMSS y PIRLS) serían una estrategia mucho más eficiente.

LECTURAS SUGERIDAS

Australian Council for Educational Research (ACER) (2019). *Minimum Proficiency Levels Described, Unpacked and Illustrated*. GAML/REF/2.

<https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/05/GAML6-REF-2-MLP-recommendations-ACER.pdf>

International Association for the Evaluation of Educational Achievement (IEA) (2017). *Measuring SDG 4: How PIRLS Can Help; How the Progress in International Reading Literacy Study (PIRLS) Helps Monitor Sustainable Development Goal 4 Targets*. Paris: UNESCO. (ED/IPS/IGE/2017/05). <https://unesdoc.unesco.org/ark:/48223/pf0000260607>

International Association for the Evaluation of Educational Achievement (IEA) (2017). M. O. Martin, I. V. S. Mullis and M. Hooper (eds), *Methods and Procedures in PIRLS 2016*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and IEA. <https://timssandpirls.bc.edu/publications/pirls/2016-methods.html>

International Association for the Evaluation of Educational Achievement (IEA) (2020). M. O. Martin, M. von Davier and I. V. S. Mullis (eds), *Methods and Procedures: TIMSS 2019 Technical Report*. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and IEA. <https://timssandpirls.bc.edu/timss2019/methods>

Khorramdel, L., L. Yin, P. Foy, J. Yoon Jung, U. Bezirhan and M. von Davier (2022). *Establishing a Concor- dance between ERCE and TIMSS/PIRLS*. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA). <https://timssandpirls.bc.edu/Rosetta-Stone-Reports/index.html>

Khorramdel, L., L. Yin, P. Foy, J. Yoon Jung, U. Bezirhan and M. von Davier (2022). *Establishing a Concor- dance between PASEC and TIMSS/PIRLS*. TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA). <https://timssandpirls.bc.edu/Rosetta-Stone-Reports/index.html>

Programme d'Analyse des Systèmes Éducatifs (PASEC) (2017). *Rapport technique de l'évaluation interna- tionale PASEC 2014*. Dakar: PASEC, Conférence des Ministres de l'Éducation des États et Gouvernements de la Francophonie (CONFEMEN). <https://www.pasec.confemen.org/wp-content/uploads/2017/12/RAP-PORT-TECHNIQUE-PASEC2014.pdf>

UNESCO (2016). *Tercer Estudio Regional Comparativo y Explicativo: Reporte Técnico*. Santiago: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000247123>

UNESCO Institute for Statistics (UIS) (2022), *Metadata for Global Indicator 4.1.1*
<https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2021/03/Minimum-Proficiency-Levels-MPLs.pdf>.

LISTA DE SIGLAS

ERCE	Estudio Regional Comparativo y Explicativo.
AIE	Asociación Internacional para la Evaluación del Logro Educativo.
ILSA	Evaluaciones Internacionales a Gran Escala.
IRT	Teoría de la Respuesta al Ítem.
MPL	Nivel Mínimo de Competencia.
PASEC	Programme d'Analyse des Systèmes Éducatifs de la CCONFEMEN.
PIRLS	Estudio Internacional para el Progreso de la Comprensión Lectora.
PV	Valor Proyectado.
SAQMEC	El Consorcio de África Meridional y Oriental para el Monitoreo de la Calidad Educativa.
DE	Desviación estándar: cuán dispersos están los datos en relación a la media.
EE	El error estándar es la desviación estándar de la media en muestras repetidas de una población.
ODS	Objetivos de Desarrollo Sostenible.
SEA-PLM	Métrica de Aprendizaje Primario del Sudeste Asiático.
TIMSS	Estudio de las Tendencias Internacionales en Matemáticas y Ciencias.
UIS	Instituto de Estadística de la UNESCO.

