

# MEDICIÓN Y SEGUIMIENTO DE LOS RESULTADOS DE APRENDIZAJE Y LAS HABILIDADES: ¿CUÁLES SON LOS DESAFÍOS HACIA EL FUTURO?

## 1. INTRODUCCIÓN

Este documento proporciona una evaluación y un análisis integral del estado y las brechas en la medición y el seguimiento de los resultados del aprendizaje y las habilidades relacionadas con el Objetivo de Desarrollo Sostenible 4 (ODS4): Educación de calidad. El ODS 4 tiene como objetivo “Garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje permanente para todos” (ONU, 2015, 2023).

Dado el enfoque de este informe en los resultados del aprendizaje y las habilidades, se explorarán con más profundidad los marcos, metodologías e indicadores relevantes relacionados con las metas 4.1, 4.4, 4.6 y 4.7 del ODS 4 y los indicadores asociados (Tabla 1). Este enfoque no niega la naturaleza subyacente de otras metas e indicadores en relación con las principales áreas de preocupación de este documento. Sin embargo, su exploración requeriría una investigación por separado.

Medir los resultados del aprendizaje y las habilidades no es trivial. La medición de los resultados del aprendizaje y las habilidades puede servir a diferentes partes interesadas, por ejemplo, padres, docentes, autoridades escolares, organismos de inspección, investigadores y formuladores de políticas. Estas partes interesadas podrían buscar información a partir de la medición de los resultados de aprendizaje y habilidades para diversos fines, incluida la elección de la escuela, la planificación de la oferta, la acreditación de estudios, la generación de teorías, la elaboración de presupuestos y reformas, y otras decisiones como recompensar, sancionar y capacitar al personal. Investigar los resultados del aprendizaje y las habilidades puede ayudar, por ejemplo, a decidir dónde es mejor gastar el presupuesto educativo, qué estrategias de enseñanza apoyan el aprendizaje de los alumnos en un contexto determinado (Clarke y Luna-Bazaldúa, 2021), y si la escolarización promueve una educación equitativa de calidad que conduzca a oportunidades de aprendizaje permanente. Dependiendo del objetivo y alcance de la medición, se pueden emplear métodos específicos de recopilación y análisis de datos. Es fundamental recordar que la Agenda 2030, de la que emanan los ODS, busca “participar plenamente en la realización de revisiones periódicas e inclusivas del progreso a nivel subnacional, nacional, regional y global” (ONU, 2015, p. 33). Esto significa que se pueden encontrar varios enfoques de medición en todo el mundo; aun así, parece haber una intención de generar datos con fines comparativos y de establecimiento de puntos de referencia.

En el contexto de una iniciativa de la magnitud de los Objetivos de Desarrollo Sostenible (ODS) y, en particular, del ODS 4 sobre Educación de Calidad, obtener una idea de las áreas en las que se han logrado avances y aquellas en las que se necesita mayor atención y mejora es una tarea crítica, aunque también desafiante. Esto es así porque el debate sobre cómo monitorear y medir los resultados y las habilidades del aprendizaje es de naturaleza continua. Iniciativas internacionales anteriores como la Educación para Todos y los Objetivos de Desarrollo del Milenio enfrentaron obstáculos en la provisión de evidencia de mejora en sus diferentes objetivos y, en particular, en las áreas menos susceptibles de mediciones cuantitativas, al incluir calificativos como *básico*, *mínimo*, *potenciador*, entre otros, haciéndolos particularmente polémicos (Torres, 1999; Unterhalter, 2014). Este informe sugiere que los ODS podrían estar sujetos a advertencias similares a las de programas anteriores, ya sea en términos de terminología o de operacionalización y, por lo tanto, para abordar las metas del ODS 4 y avanzar hacia el futuro, es relevante identificar si se genera evidencia de medición y de qué manera.

La medición del progreso en los resultados de aprendizaje y las habilidades tiene varias décadas de desarrollo y debate. Una revisión de la literatura sobre los determinantes de los resultados de la educación primaria en los países en desarrollo (Boissiere, 2004) identificó que, tradicionalmente, se han empleado los avances en psicología y sociología educativa para comprender los resultados educativos. Sin embargo, recientemente han prevalecido enfoques económicos y modelos estadísticos sofisticados. Por lo tanto, diferentes métodos –a veces antagonistas– conforman el panorama en lo que respecta a los resultados del aprendizaje y la

medición de habilidades. Predominan los enfoques económicos, incluidos los pertenecientes a las teorías de la función de producción educativa, que enfatizan el papel de una variedad de aspectos relacionados con los insumos, como el rendimiento educativo previo de los estudiantes, la educación de los padres y los ingresos, entre otros. Además, los procesos, incluida la calidad de la enseñanza, el liderazgo escolar y los resultados, generalmente presentados en puntajes de pruebas estandarizadas en un número limitado de materias, por ejemplo, lenguaje y matemáticas, han sido parte de estos enfoques (Hanushek et al., 2016; Hanushek & Rivkin, 2006; Hanushek & Woessmann, 2010; Scheerens, 1991, 1997, 2015; Scheerens et al., 2003).

**Tabla 1: Metas e indicadores del ODS 4 relacionados con los resultados del aprendizaje**

Indicador	Dominio	Definiciones requeridas	
4.1.1	Proporción de niños y adolescentes a) en grado 2 o 3, b) al final de la educación primaria y c) al final de la educación secundaria baja que han alcanzado niveles mínimos de competencia en i) lectura y ii) matemáticas, desglosada por sexo	Lectura y matemáticas	Nivel mínimo de competencia Mínimo de calidad procesal
4.2.1	Proporción de niños de 24 a 59 meses cuyo desarrollo es adecuado en cuanto a la salud, el aprendizaje y el bienestar psicosocial, desglosada por sexo	Aprendizaje, salud socioemocional.	¿Qué está bien encaminado desde el punto de vista del desarrollo?
4.4.2	Porcentaje de jóvenes y adultos que han alcanzado al menos un nivel mínimo de competencia en alfabetización digital	Habilidades de alfabetización digital	Habilidades relevantes para el empleo, el trabajo decente y el emprendimiento
4.6.1	Proporción de la población en un grupo de edad determinado que ha alcanzado al menos un nivel fijo de competencia funcional en a) alfabetización y b) aritmética, desglosada por sexo	Literatura y aritmética	Nivel fijo de aritmética funcional y alfabetización
4.7.4	Porcentaje de alumnos por grupo de edad (o nivel de educación) que demuestran una comprensión adecuada de las cuestiones relacionadas con la ciudadanía mundial y la sostenibilidad	Ciudadanía global y sostenibilidad	La definición de comprensión adecuada y lo que constituye ciudadanía mundial y sostenibilidad
4.7.5	Porcentaje de alumnos en el último grado de la educación secundaria baja que demuestran dominio del conocimiento de ciencias de la Tierra y ciencias ambientales	Ciencias ambientales y geociencia	La definición de dominio

## 1.1. OBJETIVO DEL DOCUMENTO DE POSICIÓN

El documento tiene como objetivo explorar los marcos, metodologías e indicadores existentes que se han utilizado para evaluar los resultados de aprendizaje y las habilidades, al mismo tiempo que identifica las áreas que requieren mayor atención y mejora para medir y monitorear las metas establecidas por el ODS 4. Más específicamente, el objetivo es hacer una contribución significativa a los debates e iniciativas en curso en este campo, con el fin en última instancia de establecer una comunidad de práctica internacional que pueda abordar colectivamente los desafíos futuros.

## 2. EVALUACIÓN DEL ESTADO ACTUAL DE MEDICIÓN DEL INDICADOR 4.1.1

El indicador 4.1.1 se refiere al indicador de competencia referido a tres niveles de escolaridad: primaria baja, primaria alta y secundaria baja y dos materias (lectura y matemáticas). El indicador es el siguiente:

*“4.1.1 Proporción de niños y adolescentes a) en grado 2 o 3, b) al final de la educación primaria y c) al final de la educación secundaria baja que han alcanzado niveles mínimos de competencia en i) lectura y ii) matemáticas, desglosada por sexo”.*

El formato del indicador pretende comunicar dos tipos de información:

- I. El porcentaje de estudiantes que alcanzan al menos los estándares mínimos de competencia para los dominios relevantes (matemáticas y lectura) para cada punto de medición (grados 2/3; fin de primaria y fin de secundaria baja) y
- II. Si un programa puede considerarse comparable, y las condiciones bajo las cuales el porcentaje de niños en o por encima del MPL puede considerarse comparable con el porcentaje informado por otro país.

El indicador necesita los siguientes insumos:

- **Dominio:** lectura y matemáticas. La lectura y las matemáticas se miden a nivel nacional de diversas maneras.
- **Nivel mínimo de competencia (MPL):** es el punto de referencia del conocimiento básico en un dominio (matemáticas, lectura, etc.) en una edad/grado determinado.
- **Muestra:** la muestra debe ser representativa de la población relevante.
- **Procedimientos:** los procedimientos deben cumplir con estándares mínimos de calidad.

## 2.1. DESAFÍOS

Hay algunas cuestiones críticas con respecto al reporte del indicador 4.1.1:

### *Comparabilidad de grados y niveles educativos*

El hecho de que la enseñanza primaria tenga una duración diferente en los distintos países significa que un término como "el final de la primaria" puede significar cosas distintas en lugares diferentes. Por esto, las diferencias entre los puntos de referencia de competencia y la realidad tienden a estar sistemáticamente correlacionadas con el grado dentro de los países y las regiones, complicando las comparaciones entre países y programas de evaluación, cuando el grado no es idéntico. Sin embargo, la mayoría (89%) de los países terminan su ciclo primario en los grados 5, 6 o 7, por lo que el problema podría ser menor.

### *Comparabilidad de los resultados de la evaluación a través del espacio y el tiempo*

Si bien la comparabilidad de las estadísticas entre países influye en la comparabilidad a lo largo del tiempo, esta última no implica lo primero.

- Las evaluaciones internacionales facilitan a la comparabilidad entre países, en un momento dado. Si cada programa de evaluación produce estadísticas que sean comparables a través del tiempo, las estadísticas serán comparables entre épocas y países.
- Las *evaluaciones nacionales* no son comparables entre sí en su diseño, pero aun así pueden proporcionar datos de tendencias relativamente confiables si la medición cuenta con la calidad suficiente.

### *Oportunidad e impacto político de las estadísticas*

Las evaluaciones producen estadísticas nacionales, y a menudo subnacionales, que pueden influir positivamente en la formulación y aplicación de políticas. Para que estos impactos positivos se sientan, las estadísticas no sólo deben ser precisas, sino que también deben ser ampliamente vistas como creíbles, y el tiempo de respuesta entre la evaluación y la presentación de resultados debe ser lo más corto posible.

### Calidad procesal

Las operaciones y los procedimientos sólidos y consistentes son una parte esencial de cualquier evaluación a gran escala, ya que ayudan a maximizar la calidad de los datos y minimizar el impacto de la variación de los procedimientos en los resultados. Se pueden encontrar ejemplos de estándares procesales en todas las evaluaciones internacionales de gran escala, y en muchas evaluaciones de gran escala a nivel regional, donde el objetivo es establecer coherencia procesal en todos los contextos internacionales. Muchas evaluaciones nacionales también establecen claras directrices de procedimiento para respaldar la coherencia en su puesta en práctica.

La implementación de las evaluaciones enfrenta muchas decisiones metodológicas, incluidos formatos de testeo y decisiones de muestreo. No hay necesidad de que los procedimientos y los formatos sean idénticos en todas las evaluaciones. Sin embargo, es necesario un conjunto mínimo de procedimientos (una alineación de procedimientos), de modo que se proteja la integridad de los datos y los resultados sean sólidos y razonablemente comparables para cualquier país determinado a lo largo del tiempo, pero también entre países en un momento determinado.

### Costos financieros de las evaluaciones para los países

Las evaluaciones son relativamente costosas en comparación con otros sistemas de recopilación de datos como EMIS. Sin embargo, incluso para los países en desarrollo, el costo de evaluar los resultados sistemáticamente es extremadamente bajo en relación con el costo global de la escolarización y en relación con el costo de no hacer mediciones<sup>1</sup>.

### Baja cobertura de evaluaciones internacionales, especialmente en países de ingresos bajos y medio bajos.

El indicador 4.1.1 de los ODS se informa utilizando varios estudios transversales que son internacionales ([PIRLS](#), [TIMSS](#)) o regionales ([PILNA](#), [SEA-PLM](#), [PASEC](#), [LLECE](#), [SACMEQ](#)) (Tabla 2). Estas herramientas no han sido diseñadas para informar sobre los ODS pero, en 2018, la Alianza Mundial para Monitorear el Aprendizaje ([GAML](#)) y el Grupo de Cooperación Técnica sobre los indicadores del ODS 4 ([TCG](#)) acordaron que estas evaluaciones podrían utilizarse para informar sobre el aprendizaje en función de los niveles de competencia que mejor se ajusten al MPL global.

**Tabla 2: Programas de evaluación por grado o edad y uso para reportar el indicador 4.1.1 del ODS**

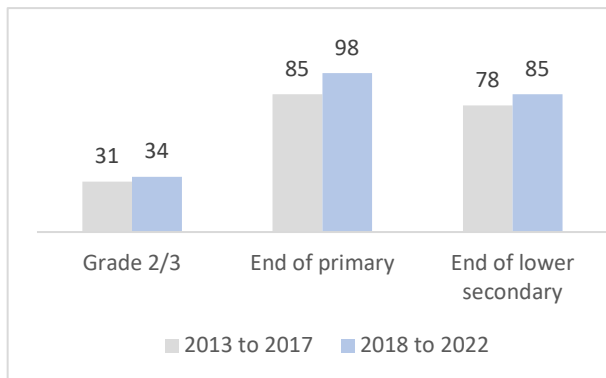
Grado	Programa de evaluación internacional
Basado en la escuela	
<b>ODS 4.1.1a: Primeros grados</b>	
2	<a href="#">EGMA</a> , <a href="#">EGRA</a> , <a href="#">PASEC</a>
3	<a href="#">EGMA</a> , <a href="#">EGRA</a> , <a href="#">ERCE</a> , <a href="#">AMPLa</a>
<b>ODS 4.1.1b: Fin de la primaria</b>	
4	<a href="#">PILNA</a> , <a href="#">LaNA</a> , <a href="#">PIRLS</a> , <a href="#">TIMSS</a>
5	<a href="#">SEA-PLM</a>
6	<a href="#">LaNA</a> , <a href="#">PASEC</a> , <a href="#">PILNA</a> , <a href="#">SACMEQ</a> , <a href="#">ERCE</a> , <a href="#">AMPLb</a>
<b>ODS 4.1.1c: Fin de la educación secundaria baja</b>	
8	<a href="#">TIMSS</a>
Edad/ 15 años	<a href="#">PISA</a>

<sup>1</sup>Para obtener más información sobre el cálculo de costos, consulte este documento: [https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2023/05/Countrys-reporting-option\\_Zambia\\_2023.05.15\\_FINAL.pdf](https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2023/05/Countrys-reporting-option_Zambia_2023.05.15_FINAL.pdf)

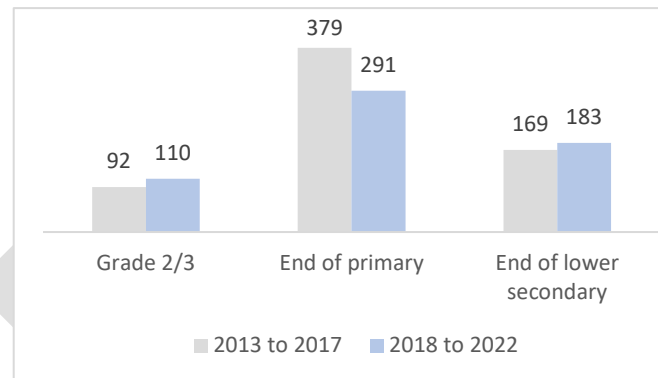
Sin embargo, la producción de resultados de aprendizaje comparables no avanza lo suficientemente rápido ni de manera equitativa. Independientemente del criterio de cobertura (número de países o población), la cobertura es mucho mayor al final de la primaria y del primer ciclo de secundaria que en los grados 2 o 3 (Figura 1).

**Figura 1 - Cobertura de las evaluaciones de aprendizaje, por nivel de educación**

a. Número de países



b. Población en edad escolar, en millones



### *Evaluaciones nacionales: limitación de alineación e incertidumbre sobre la calidad de los procedimientos*

Si bien los datos de muchas evaluaciones nacionales de aprendizaje están disponibles, cada país establece sus propios estándares, lo que lleva a definiciones inconsistentes de los niveles de desempeño. Por lo tanto, el análisis de los resultados queda limitado a una prueba, una metodología y una escala.

Aunque las metodologías tienden a converger entre las evaluaciones internacionales y regionales, sigue siendo difícil situar a las evaluaciones en una evaluación nacional común de nivel de referencia, y además existe incertidumbre con respecto al conjunto de procedimientos utilizados para el muestreo, la gestión de datos y la elaboración de informes.

Las operaciones y los procedimientos sólidos y consistentes son una parte esencial de cualquier evaluación a gran escala, ya que ayudan a maximizar la calidad de los datos y minimizar el impacto de la variación de los procedimientos en los resultados. Se pueden encontrar ejemplos de estándares procesales en todas las evaluaciones internacionales de gran escala, y en muchas evaluaciones de gran escala a nivel regional, donde el objetivo es establecer coherencia procesal en todos los contextos internacionales. Muchas evaluaciones nacionales también establecen claras directrices de procedimiento para respaldar la coherencia en su puesta en práctica.

La implementación de las evaluaciones enfrenta muchas decisiones metodológicas, incluidos formatos de testeo y decisiones de muestreo. No hay necesidad de que los procedimientos y los formatos sean idénticos en todas las evaluaciones. Sin embargo, es necesario un conjunto mínimo de procedimientos (una alineación de procedimientos), de modo que se proteja la integridad de los datos y los resultados sean sólidos y razonablemente comparables para cualquier país determinado a lo largo del tiempo, pero también entre países en un momento determinado.

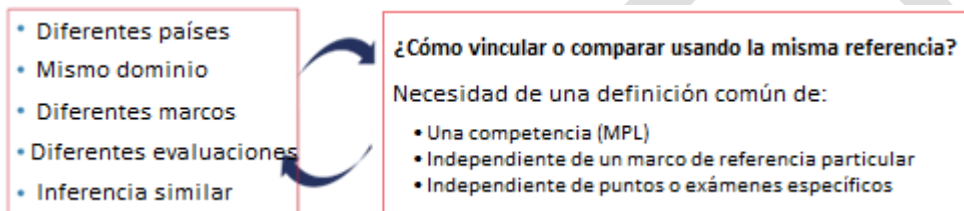
## 2.2. AVANCES Y DESAFÍOS PENDIENTES

Lograr estadísticas que sean comparables entre programas y países es quizás más difícil de lo que se supone. Esto se debe a que las diferentes regiones tienen distintas tradiciones en cuanto a la rigurosidad de los puntos de referencia de competencia en los diferentes grados. Además, estas realidades complican aún más las comparaciones entre países, que a menudo implican comparar grados ligeramente diferentes, incluso en el mismo nivel educativo. El foco del trabajo del UIS ha sido la definición del nivel mínimo de competencia como el concepto que permite la presentación de reportes, y un conjunto de estrategias de vinculación al marco de competencia.

Para producir datos comparables en todos los programas al nivel mínimo de competencia, el UIS ha decidido alinearse a un concepto en lugar de a un puntaje basado en puntos, estrategia que requiere un proceso de alineación para expresar todas las evaluaciones en ese punto de referencia resumido en la Figura 2.

La tarea implicó la creación de un conjunto de estándares globales, lo cual fue un proceso lento pero necesario. Esto permitiría que evaluaciones dispares hicieran referencia a normas, evitando tener que realizar la misma evaluación. Esto sería posible con el siguiente [conjunto de herramientas](#) :

**Figura 2 – Vincular o comparar usando la misma referencia**



## 2.3. ESTÁNDARES

### 2.3.1. EL NIVEL MÍNIMO DE COMPETENCIA (MPL)

El nivel mínimo de competencia (MPL) es el punto de referencia del conocimiento básico en un dominio (matemáticas, lectura, etc.) en una edad/grado determinado medido a través de evaluaciones de aprendizaje. El MPL es un punto de referencia para informar sobre las competencias mínimas en cada nivel de escolaridad, sin la necesidad de requerir una prueba para lograr la comparabilidad.

El Descriptor de Nivel de Competencia (PLD) del MPL es el estándar clave para cada grado y dominio, ya que permite el uso de múltiples evaluaciones para informar el indicador. La Tabla 3 presenta las definiciones globales de MPL para los dominios de matemáticas y lectura.<sup>2</sup>

El primer paso en la implementación del MPL fue acordar con los programas de evaluación internacionales el PLD en cada programa de evaluación que esté mejor alineado con el PLD de la MPL. Este primer paso acordado en 2018 necesita más pasos de validación por medio de un ejercicio de establecimiento de estándares, ya que los programas de evaluación no han sido diseñados para medir el ODS 4, y luego son utilizados para:

- aproximar los grados de interés que se van a informar;
- utilizar los niveles de competencia existentes en el programa de evaluación;
- identificar el PLD que esté mejor alineado con los niveles mínimos de competencia;
- **utilizar ese nivel para informar** hasta que finalice el ejercicio de establecimiento de normas;

<sup>2</sup>Fue acordado reportarlo de acuerdo con la definición textual del MPL para cada dominio y nivel en las Evaluaciones Internacionales (CNA). Esto fue establecido mediante la realización de un análisis de los descriptores de nivel de competencia (PLD) de evaluaciones internacionales, regionales y dirigidas por la comunidad, en lectura y matemáticas.

- desarrollar la capacidad técnica nacional pero no abordar directamente el desarrollo de evaluaciones nacionales.

**Tabla 3. Niveles mínimos de competencia en lectura y matemáticas - Indicador 4.1.1**

Nivel educativo	Descriptor	
	Lectura	Matemáticas
Grado 2	Leen y comprenden la mayoría de las palabras escritas, especialmente las que les resultan familiares, y extraen información explícita de las oraciones.	Los estudiantes demuestran habilidades en sentido numérico y computación, reconocimiento de formas y orientación espacial.
Grado 3	Los estudiantes leen en voz alta palabras escritas con precisión y fluidez. Entienden el significado general de frases y textos breves.	
Grado 4 -6	Los estudiantes interpretan y dan algunas explicaciones sobre las ideas principales y secundarias en diferentes tipos de textos. Establecen conexiones entre las ideas principales de un texto y sus experiencias personales, así como conocimientos generales.	Los estudiantes demuestran habilidades en sentido numérico y cálculo, medición básica, lectura, interpretación y construcción de gráficos, orientación espacial y patrones numéricos.
Grado 8 y 9	Los estudiantes establecen conexiones entre las ideas principales de diferentes tipos de texto y las intenciones del autor. Reflexionan y sacan conclusiones a partir del texto.	Los estudiantes demuestran habilidades en computación, problemas de aplicación, comparación de tablas y gráficos y uso de representaciones algebraicas.

Fuente: Instituto de Estadística de la UNESCO, 2021, p. 4

### 2.3.2. EL MARCO DE COMPETENCIA GLOBAL (GPF)

El Marco de Competencia Global es una referencia global útil que define los niveles de competencia en lectura y matemáticas que se espera que los alumnos demuestren al final de cada nivel de grado, desde el grado 1 al 9, como referencia común. Los cuatro niveles descritos en el GPF (Inferior a Cumple parcialmente, Cumple parcialmente, Cumple y Supera la competencia mínima global) forman una escala común de logro bajo a alto.

Además, se ha creado un estándar adicional. El marco de Competencia Global y sus niveles de competencia relacionados brindan orientación sobre el conjunto mínimo de habilidades que los estudiantes deben adquirir en el camino hacia el dominio de la lectura y las matemáticas.

### 2.3.3. UN CONJUNTO DE HERRAMIENTAS PARA VINCULAR LOS PROGRAMAS DE EVALUACIÓN AL MPL

La vinculación de una evaluación nacional, regional o internacional a la definición del nivel de competencia global representada por el MPL requiere una metodología para identificar los mismos conceptos/definiciones en programas de evaluación creados para propósitos completamente diferentes y poder expresarlos de una manera que permita cierto grado de comparabilidad que, a su vez, logre inferencias justas sobre los sujetos (países) comparados.

El proceso de hacer comparables esas diferentes evaluaciones, llamado “moderación”, podría basarse en una calibración estadística o no estadística.

Se han sugerido diferentes soluciones para obtener datos que puedan usarse para medir y monitorear el ODS 4.1 y sus indicadores. Tres de los más discutidos y apoyados por la comunidad de evaluación internacional son el proyecto Piedra Rosetta, la Metodología de Vinculación de Políticas y las Evaluaciones de los Niveles Mínimos de Competencia (AMPL). A continuación, proporcionamos información básica sobre cada una de



estas iniciativas, así como sus principales fortalezas y limitaciones. Esta información se incorporará a la sección final de este documento, que será una propuesta de agenda a seguir por la comunidad internacional de medición educativa en el corto y mediano plazo. Consulte el [documento](#) sobre las opciones que tienen los países para informar el indicador 4.1.1 para obtener más información sobre estas diferentes estrategias de vinculación y un resumen de costos, beneficios, estado de ejecución, hitos ejecutados y pendientes, y marco temporal.

### *Piedra Rosetta*

Una de las iniciativas que se ha probado para intentar armonizar los datos de diferentes evaluaciones es el proyecto **Piedra Rosetta** que fue liderado por la Asociación Internacional para la Evaluación del Rendimiento Educativo (IEA). El Estudio Piedra Rosetta está diseñado para medir el progreso global hacia el ODS 4.1.1 relacionando las evaluaciones de aprendizaje nacionales y regionales con las evaluaciones de aprendizaje internacionales. Debe su nombre al famoso descubrimiento arqueológico que permitió la traducción entre diferentes lenguas escritas: la Piedra Rosetta. El objetivo del estudio es proporcionar a los países que participaron en evaluaciones regionales o nacionales, pero no en evaluaciones internacionales, información sobre la proporción de niños de escuela primaria que han alcanzado un nivel mínimo de competencia en alfabetización y aritmética (ODS 4.1.1) que les permita comparaciones internacionales.

En un primer esfuerzo para implementar este enfoque y establecer tablas de concordancia, las evaluaciones regionales Estudio Regional Comparativo y Explicativo (ERCE) y Programa para el Análisis de Sistemas Educativos de CONFEMEN ([PASEC](#)) están vinculadas a dos evaluaciones internacionales de la IEA, a saber, Estudio Internacional de Tendencias en Matemáticas y Ciencias ([TIMSS](#)) para matemáticas y Estudio Internacional de Progreso en Comprensión Lectora ([PIRLS](#)) para lectura.

### *Vinculación de políticas (Policy Linking)*

Otro enfoque relacionado con la armonización de evaluaciones es la metodología **Policy Linking** (Instituto de Estadística de la UNESCO, 2021) que es un método no estadístico que utiliza el juicio para alinear y hacer coincidir elementos de la evaluación nacional con el Marco de Competencia Global (GPF). Este proceso establece los puntos de referencia globales comparables internacionalmente basándose en los descriptores de cada punto de referencia especificado en el GPF.

Se deben realizar tres tareas principales (alineación, coincidencia y establecimiento de puntos de referencia) en un taller de 5 a 6 días con 15 a 20 panelistas (docentes), expertos en los planes de estudio y en las evaluaciones de cada grado/materia, para identificar y establecer, si es posible, los puntos de referencia requeridos para la presentación de informes internacionales sobre el indicador 4.1.1 de los ODS.

Para producir puntos de referencia confiables para la presentación de informes internacionales, el Policy Linking Toolkit (PLT) especifica cinco criterios que incluyen: un número suficiente de elementos nacionales están alineados con el GPF; las muestras son representativas a nivel nacional; y las evaluaciones nacionales se administran de acuerdo con estándares mínimos de calidad. Si no se cumplen los criterios, el taller se considerará una actividad de capacitación.

La metodología de Policy Linking se propuso durante la reunión de la Alianza Mundial para Monitorear el Aprendizaje (GAML) en 2017, se puso a prueba en 2019 por primera vez, se revisó en 2020 y se puso a prueba nuevamente en 2021-2022. Posteriormente, el PLT se revisó en enero de 2023 y se encuentra en fase piloto.

### *Evaluaciones de los Niveles Mínimos de Competencia (AMPL)*

Un tercer método que se ha propuesto como solución son las Evaluaciones de los Niveles Mínimos de Competencia (AMPL). Los AMPL son herramientas robustas destinadas a medir el logro de un único nivel de competencia para cada uno de los dominios de lectura y matemáticas en un nivel determinado del ciclo

educativo. Las herramientas AMPL permiten identificar la proporción de niños y jóvenes estudiantes en cada nivel de educación que están alcanzando al menos el Nivel Mínimo de Competencia (MPL). Esto permite a los países producir datos internacionales comparables sobre resultados de aprendizaje para informar sobre el indicador global ODS 4.1.1.

AMPL-b es el primer AMPL desarrollado en 2021 tanto en inglés como en francés y se implementó en seis países africanos como parte del proyecto MILO en 2021: Burkina Faso, Burundi, Costa de Marfil, Kenia, Senegal y Zambia. AMPL-b se administró como módulo independiente en Sierra Leona en 2022 y está previsto que se implemente en Jordania y Pakistán.

AMPL-a, que mide las competencias en los primeros grados, está en desarrollo y se pondrá a prueba y se administrará en 2023 en ambos idiomas, inglés y francés.

#### *2.3.4. PAÍSES: ALTERNATIVAS PARA LA PRESENTACIÓN DE INFORMES.*

Para guiar la elección de cómo medir el aprendizaje y garantizar que los datos de la evaluación sean consistentes con los objetivos estratégicos a largo plazo de una toma de decisiones efectiva, el UIS, la UNESCO, el Banco Mundial y UNICEF han desarrollado un conjunto de principios en los que se basa esta sección, conocidos como el Pacto para los Datos sobre el Aprendizaje.

Un conjunto de principios es importante no sólo para diseñar evaluaciones o decidir qué evaluación adquirir "de la estantería", sino también para desarrollar un sistema de evaluación para el país de cada uno. Esos principios deben basarse en lo que ya existe; permitir flexibilidad para garantizar la alineación con las necesidades de los países (no una solución única para todos); fomentar la apropiación nacional mediante un enfoque impulsado por la demanda; garantizar que los datos sean relevantes para la toma de decisiones.

Un sistema de evaluación nacional debería ser bueno no sólo para informar sino también para gestionar mejoras en todos los niveles de educación, para desarrollar la capacidad de toma de decisiones y para vincular las evaluaciones a nivel de sistema con las evaluaciones formativas y las prácticas en el aula. Para garantizar que las evaluaciones puedan monitorear con precisión el progreso para la toma de decisiones, los datos también deben ser comparables internacionalmente. Cada país debería tener una evaluación que, de una forma u otra, haya sido diseñada o pueda utilizarse para la comparabilidad internacional —un compromiso en el proceso de los ODS.

Las opciones de los países para informar se reflejan en la tabla 4, pero la elección debe guiarse en base a qué evaluaciones son adecuadas para su propósito y son las más rentables para ellos, teniendo en cuenta la situación inicial del país, el objetivo de tener comparabilidad a lo largo del tiempo y la representatividad de los resultados a nivel nacional.

Cabe señalar dos casos especiales. Se podrían utilizar evaluaciones nacionales para informar, sujeto a la implementación de vinculaciones estadísticas que podrían implementarse utilizando módulos calibrados como AMPL. Otras herramientas, como los niveles mínimos de competencia (MPL) y el Marco de Competencia Global, sirven para comprender y comparar los estándares globales, mientras que el Policy Linking (PL) sirve para involucrar a las partes interesadas nacionales y analizar la evaluación con respecto a esos estándares. El UIS considera que a los países les gustaría informar globalmente sobre el indicador 4.1.1: agregan un módulo calibrado a la evaluación nacional, como AMPL, y lo complementan con el Policy Linking (descrito anteriormente) para propósitos de capacitación, dado que la metodología aún está en una fase piloto.

Un segundo caso está relacionado con los primeros grados, o 4.1.1a, donde la existencia de herramientas cobra especial relevancia - como la Evaluación de Lectura/Matemáticas en los primeros grados (EGRA/EGMA), las evaluaciones dirigidas por ciudadanos de PAL Network y el Módulo de Competencias Fundacionales de UNICEF, de las encuestas de hogares MICS – ya que potencialmente podrían servir como

reportes. Aunque estas evaluaciones se han aplicado globalmente, actualmente no se pueden utilizar para reportes globales, principalmente porque no estaban destinadas a generar datos comparables. Sin embargo, tienen el potencial para usarse como tales, y el UIS está estudiando cómo aprovechar al máximo dichas evaluaciones. Para obtener más detalles, consulte este blog [aquí](#).

**Tabla 4: Alternativas para que los países informen el indicador 4.1.1 de los ODS**

	4.1.1.a	4.1.1.b	4.1.1.c	Cobertura
<b>Evaluaciones nacionales: vinculación estadística por medio de módulos calibrados</b>				
AMPL	•	•		
Módulo PISA			•	
<b>Participar en una evaluación internacional</b>				
PILNA		•		Islas del pacífico
PASEC	•	•	•	África (francófona) principalmente
SACMEQ		•		África (sur y este)
SEA-PLM		•		Sudeste asiático
LLECE	•	•		América Latina
TIMSS	•	•	•	Global
PIRLES	•	•		Global
PISA			•	Global

### 3. EVALUACIÓN DEL ESTADO ACTUAL DE MEDICIÓN DE LOS INDICADORES 4.4.2, 4.6.1, 4.7.4 Y 4.7.5

Se han logrado avances significativos en el establecimiento de marcos metodológicos para estos indicadores. El desarrollo de estos marcos proporciona un enfoque estructurado, ofreciendo directrices y principios para la recopilación y el análisis de datos. Sin embargo, a pesar de los marcos metodológicos definidos, existe un problema sustancial en forma de baja cobertura de datos, particularmente en los países de ingresos bajos y medios (PIBM). Abordar este desafío requiere un esfuerzo concertado para establecer definiciones y métricas comunes, asegurando un enfoque estandarizado en todas las evaluaciones (Tabla 5).

**Tabla 5: Principales parámetros de disponibilidad de evaluación**

Indicadores	Marco metodológico	Fuentes de datos	Administración. último	Longitud del ciclo	Cobertura de países	Cobertura de población (%)
4.4.2	<a href="#">Sí</a>	Evaluación de competencias de adultos (PIAAC)	2017	indefinido	6	7
4.6.1	<a href="#">Sí</a>				5	3
4.7.4	<a href="#">Sí</a>	ICCS	2016	6/7 años	23	0
4.7.5	<a href="#">Sí</a>	TIMSS, PISA	2019/2022	4/5 años	25	12

#### 3.1. ASUNTOS PENDIENTES

El hecho es que los países de ingresos bajos y medios a menudo carecen de los recursos y la infraestructura necesarios para desarrollar e implementar herramientas para medir estos indicadores de manera efectiva. Sin suficientes datos sobre estos indicadores, a estos países les resulta difícil identificar áreas específicas de mejora y asignar recursos de manera eficiente. Además, esta brecha de datos inhibe los esfuerzos internacionales para brindar apoyo específico a los países que más lo necesitan, lo que obstaculiza el progreso

global hacia el logro del ODS 4. Para abordar este problema se requiere no solo el desarrollo de herramientas de medición apropiadas, sino también iniciativas específicas de creación de capacidad en los países de ingresos bajos y medios para garantizar que estos indicadores se midan de manera integral y precisa, proporcionando una base para la toma de decisiones informadas y la formulación de políticas en el ámbito de la educación.

#### 4. METODOLOGÍA PARA ARMONIZAR LOS CUESTIONARIOS DE CONTEXTO

Más allá de armonizar datos sobre resultados de aprendizaje y habilidades, armonizar cuestionarios de contexto de diferentes evaluaciones a gran escala es de vital importancia para permitir análisis comparativos sólidos de tendencias, patrones y determinantes de la desigualdad educativa entre países y a lo largo del tiempo. Esta armonización de datos permitiría a los investigadores reunir grandes conjuntos de datos longitudinales que podrían proporcionar conocimientos novedosos sobre cuestiones clave como la desigualdad en el aprendizaje, la segregación escolar, la privatización y el desempeño en los países de ingresos bajos y medios. Al colaborar entre instituciones y aprovechar las evaluaciones existentes de diversos contextos, los cuestionarios armonizados podrían desbloquear oportunidades para una importante investigación que sea altamente relevante para las políticas y contribuya directamente al seguimiento del progreso en diferentes indicadores del ODS 4.

Si bien el objetivo principal de las CNA es estimar medidas de resultados de aprendizaje para un país, también recopilan un rico conjunto de información básica sobre docentes, estudiantes y escuelas. De los estudiantes, las CNA generalmente recopilan información sobre su experiencia en las escuelas, sus actitudes hacia las materias que se imparten y las características de sus padres y hogares, además de información demográfica básica sobre edad y sexo (Tabla 6). De los docentes, las CNA recopilan información sobre sus actitudes hacia la enseñanza, sus opiniones sobre los recursos didácticos, su formación académica y su desarrollo profesional continuo; y de las escuelas, las CNA recopilan información sobre la infraestructura, la ubicación y las opiniones de los directores de las escuelas sobre la disponibilidad de recursos y cómo interactúan con los padres. En el caso de las escuelas, existe cierta variación en la objetividad de los datos recopilados.

**Tabla 6: Cuestionarios típicos y datos recopilados en las CNA**

Prueba cognitiva	Elementos de la prueba (preguntas) para medir los resultados de aprendizaje
Cuestionario para estudiantes	<p>Información demográfica básica (sexo, edad)</p> <p>Hogar y entorno socioeconómico</p> <p>Experiencias relacionadas con la escuela (incluida la exposición al acoso)</p> <p>Experiencias relacionadas con el aprendizaje (actividades en el aula)</p> <p>Autopercepciones, intereses y aspiraciones relacionadas con diferentes temas.</p> <p>Uso y dominio de las TIC</p>
Cuestionario para docentes	<p>Información demográfica y de antecedentes (sexo, edad, años de docencia, materias impartidas)</p> <p>Cualificaciones y formación</p> <p>Tipos de prácticas docentes utilizadas y desafíos enfrentados</p>
Cuestionario del director de escuela	<p>Información demográfica y de antecedentes (sexo, edad, años de experiencia)</p> <p>Cualificaciones y educación</p> <p>Características de la escuela</p> <p>Opiniones sobre disponibilidad y adecuación de recursos</p> <p>Gestión y gobernanza</p> <p>Interacción con padres y comunidades escolares.</p> <p>Desafíos enfrentados en la enseñanza.</p>

Es a partir de estos cuestionarios que se puede determinar si un indicador de los ODS puede estimarse o no. Son también estos cuestionarios los que determinan para qué subpoblaciones se pueden estimar los indicadores y cómo informan la dimensión de equidad. Generalmente, estos incluyen la ubicación urbana y rural de la escuela, el nivel socioeconómico (NSE) del estudiante (en relación con otros estudiantes, no con la población) y el sexo del estudiante (o docente)<sup>3</sup>.

#### 4.1. CUESTIONARIOS DE CONTEXTO - ARMONIZACIÓN

Interpretar y obtener asesoramiento sobre políticas plantea desafíos importantes debido a las diferencias inherentes en las definiciones en varias dimensiones, como las clasificaciones rurales/urbanas, el estatus (o riqueza) socioeconómico, el período de referencia, y más. Las disparidades en estas definiciones dificultan la comparabilidad de los resultados de las evaluaciones y la interpretación de los datos: por ejemplo, lo que constituye "rural" en un país puede diferir de otro. De manera similar, las distintas definiciones de NSE pueden afectar el análisis de las disparidades en los resultados educativos entre diferentes estratos sociales o económicos. Cerrar estas brechas de definición requiere colaboración internacional y el desarrollo de marcos estandarizados que garanticen la uniformidad en las definiciones, permitiendo evaluaciones precisas y facilitando comparaciones significativas entre regiones y contextos socioeconómicos.

La armonización de los cuestionarios de contexto también puede servir para otros propósitos valiosos, entre ellos:

- *Comparar investigaciones:* Los cuestionarios armonizados permitirían a los investigadores comparar contextos educativos entre países, regiones o a lo largo del tiempo. Esto puede ayudar a identificar tendencias, similitudes y diferencias en los sistemas, políticas y prácticas educativas.
- *Análisis de políticas:* Los formuladores de políticas podrían utilizar datos armonizados para evaluar la efectividad de las políticas e intervenciones educativas comparando los resultados en diferentes contextos. Esto ayuda a la formulación de políticas basadas en evidencia.

<sup>3</sup>Para obtener más información, consulte este [documento](#) del UIS: 'Monitoring of the Sustainable Development Goals using Large-Scale International Assessments' (2022).

- *Análisis de equidad:* Los investigadores podrían utilizar datos armonizados para investigar las desigualdades educativas, incluidas las disparidades en el acceso a recursos y oportunidades. Esto podría informar los esfuerzos para reducir las desigualdades educativas.
- *Desarrollo curricular:* Los datos contextuales armonizados pueden ayudar en el desarrollo de planes de estudio que estén mejor alineados con las necesidades y desafíos de los estudiantes en diferentes contextos.
- *Formación docente:* Comprender los factores contextuales que afectan la enseñanza y el aprendizaje puede servir para los programas de formación docente, garantizando que los educadores estén bien preparados para los desafíos específicos que puedan enfrentar.
- *Asignación de recursos:* Los gobiernos pueden utilizar datos armonizados para asignar recursos educativos de manera más efectiva, centrándose en las áreas con mayor necesidad.
- *Investigación intercultural:* los investigadores pueden realizar estudios interculturales para explorar cómo los factores culturales impactan en los resultados de la educación y el aprendizaje.

## 4.2. DESAFÍOS EN LA ARMONIZACIÓN

Los desafíos se describen brevemente y todos exigen una adaptación para que sean relevantes y sensibles al contexto:

1. *Diferencias culturales y lingüísticas:* Las variaciones en los factores culturales y lingüísticos pueden conducir a diversas interpretaciones de las preguntas y respuestas existentes, requiriendo una adaptación cultural.
2. *Especificidad contextual:* Los contextos educativos varían ampliamente entre países y regiones.
3. *Sistemas educativos diferentes:* Las diferencias en los sistemas y políticas educativas entre países exigen la armonización de cuestiones que originalmente no fueron diseñadas para ser aplicables universalmente.
4. *Métodos de recopilación de datos:* Los países pueden emplear diferentes métodos, procedimientos e instrumentos para la recopilación de datos. Es complejo armonizarlos en los cuestionarios existentes manteniendo al mismo tiempo la calidad de los datos.
5. *Variabilidad de las respuestas:* Las respuestas de los individuos a las preguntas existentes pueden variar según las normas y expectativas culturales. Es importante que existan patrones coherentes de interpretación y respuesta durante el proceso de armonización.
6. *Consideraciones éticas y de privacidad:* La recopilación de datos existente a menudo involucra información confidencial sobre estudiantes, maestros y escuelas. Garantizar la privacidad de los datos y adherirse a directrices éticas mientras se armonizan los datos puede ser una tarea compleja.
7. *Control de calidad:* Mantener la calidad y coherencia de los datos en diversos contextos al armonizar los cuestionarios existentes requiere medidas rigurosas de control de calidad, que pueden consumir muchos recursos.
8. *Contextos cambiantes:* Los contextos educativos evolucionan con el tiempo y los cuestionarios de contexto existentes pueden quedar obsoletos, lo que requiere regularidad para mantener los cuestionarios relevantes durante la armonización.
9. *Sensibilidades políticas y culturales:* Algunas preguntas de los cuestionarios existentes pueden abordar cuestiones políticas o culturales delicadas que exigen un terreno común y un acuerdo sobre la redacción adecuada.
10. *Estandarización de datos:* Armonizar los formatos de datos y los esquemas de codificación puede ser un desafío técnico en comparación con los cuestionarios existentes, especialmente cuando los países utilizan sistemas diferentes.

## 5. ESTABLECIENDO UNA AGENDA PARA LA COMUNIDAD DE MEDICIÓN

### 5.1. UN PLAN PARA GUIAR LA PRESENTACIÓN DE INFORMES DE LOS PAÍSES Y GARANTIZAR LA CALIDAD Y LA ALINEACIÓN – 4.1.1

El protocolo existente para la presentación de informes y el menú de opciones, incluidas las formas alternativas de presentación de informes y menú de opciones disponibles, para garantizar que los usuarios sean plenamente conscientes de las propiedades y limitaciones de los datos.

Si bien iniciativas como Piedra Rosetta, Policy Linking y AMPL han trabajado para armonizar diferentes evaluaciones educativas, todavía se necesita un plan estandarizado para evaluar sistemáticamente qué evaluaciones son adecuadas para incluir en estos esfuerzos de armonización. A medida que surgen más evaluaciones nacionales e internacionales a gran escala, es fundamental contar con criterios claros para analizar su calidad, comparabilidad y viabilidad para la armonización.

Un plan integral debe describir los factores clave a examinar para cada evaluación bajo consideración. Por ejemplo:

- Alineación con los estándares y marcos de aprendizaje - las evaluaciones deben medir adecuadamente el plan de estudios y las habilidades previstos.
- Propiedades psicométricas - evidencia de confiabilidad, validez, dificultad apropiada, discriminación, etc.
- Representatividad - las muestras deben reflejar las poblaciones objetivo.
- Comparabilidad de las administraciones - Procedimientos de administración coherentes y estandarizados.
- Transparencia de los procesos - el diseño de la evaluación, el muestreo y el análisis deben estar bien documentados.
- Capacidad de vinculación - suficientes elementos/niveles de competencia equivalentes para permitir la vinculación.
- Participación de las partes interesadas - inclusión de expertos durante todo el diseño y la implementación.
- Viabilidad de la participación - Costos, cronogramas y cargas razonables para los países.

En un anexo de este documento de posición se desarrollará un plan detallado que incorpora estos elementos y representará el primer punto de la agenda que se propondrá a la comunidad internacional de evaluación educativa. Este plan permitirá una investigación rigurosa de las evaluaciones para determinar su idoneidad y capacidad técnica para las iniciativas de armonización. De esta manera se podrán fortalecer los estándares y la participación globales. Por ejemplo, se debe intentar desarrollar un modelo acordado a través de la red GAML.

### 5.2. ARMONIZACIÓN DE CUESTIONARIOS DE CONTEXTO

Armonizar los cuestionarios de contexto existentes en evaluaciones internacionales a gran escala requiere una cuidadosa consideración de estos desafíos para garantizar que los datos armonizados resultantes sean confiables y comparables en diferentes contextos educativos.

La estrategia principal sería la creación y el mantenimiento sistemático de un conjunto de datos armonizados de evaluaciones internacionales a gran escala en educación, que proporcione indicadores longitudinales de niveles y tendencias en el logro educativo (ver sección anterior) y sus posibles determinantes, a nivel de país. Este conjunto de datos armonizados incluiría indicadores de evaluaciones escolares nacionales, globales y regionales a gran escala que cumplan con los requisitos mínimos de calidad establecidos por un plan creado

para este propósito (ver la siguiente sección). La idea sería garantizar que en la armonización sólo se incluyan evaluaciones que contengan información sobre los antecedentes de los estudiantes, las características de las escuelas y los resultados de aprendizaje, que sea comparable a lo largo del tiempo y entre países. Estudios anteriores han desarrollado metodologías de armonización (Angrist et al., 2021; Gust et al., 2022; Patel & Sandefur, 2019; Sandoval-Hernandez, 2022). La armonización de los cuestionarios de contexto debería basarse en este trabajo mediante la creación de indicadores para los antecedentes de los estudiantes, los docentes y los directores, los recursos escolares y la desigualdad educativa.

### 5.3. SOLUCIONES PROPUESTAS PARA MEJORAR LA COBERTURA - 4.6.1

- Consideración de alternativas para reemplazar el indicador PIAAC, como la tasa de alfabetización que tiene un alto índice de cobertura y frecuencia en los informes.
- Utilización de inteligencia artificial para producir indicadores para la población adulta (por ejemplo, leer y calificar cuerpos de texto disponibles).