



**ESTADO  
DE LA EDUCACIÓN**

**Informe Estado de la Educación 2025**

## **Investigación**

Pobreza de las evaluaciones  
educativas: El caso de las pruebas  
estandarizadas sumativas de sexto  
grado y último año de secundaria  
en Español y Matemática aplicadas  
en Costa Rica en el 2023

**Investigadores:**

Eiliana Montero-Rojas  
Katherine Barquero Mejías  
Ricardo Poveda Vásquez  
Marielos Murillo Rojas

San José | 2025



370  
P739p

Pobreza de las evaluaciones educativas: El caso de las pruebas estandarizadas sumativas de sexto grado y último año de secundaria en Español y Matemática aplicadas en Costa Rica en el 2023 / Eiliana Montero-Rojas...[et al.]. -- San José, C.R. : CONARE - PEN, 2025.

1 recurso en línea (78 páginas): archivos de texto PDF, 1.550 KB

ISBN 978-9930-654-24-8

Investigación para el Informe Estado de la Educación 2025

1. EDUCACIÓN PRIMARIA. 2. EDUCACIÓN SECUNDARIA. 3. PRUEBAS ESTANDARIZADAS. 4. POLÍTICA EDUCATIVA. 5. CALIDAD DE LA EDUCACIÓN. 6. COSTA RICA. I. Montero-Rojas, Eiliana. II. Barquero Mejías, Katherine. III. Poveda Vásquez, Ricardo. IV. Murillo Rojas, Marielos Título.



### Información de los autores

**Eiliana Montero-Rojas.** <https://orcid.org/0000-0002-6770-792X>

**Katherine Barquero Mejías.** <https://orcid.org/0000-0003-1760-9026>

**Ricardo Poveda Vásquez.** <https://orcid.org/0000-0001-5052-0117>

**Marielos Murillo Rojas.** <https://orcid.org/0009-0009-1629-9485>

Esta obra se comparte bajo la licencia

**Reconocimiento – No Comercial – Compartir Igual**

**(CC-BY-NC-SA)**



Permite usar una obra para crear otra obra o contenido, modificando o no la obra original, siempre que se cite al autor, la obra resultante se comparta bajo el mismo tipo de licencia y no tenga fines comerciales

## **Contenido**

Descargo de responsabilidad .....	4
Introducción.....	4
Pobreza de los aprendizajes y pobreza de las evaluaciones: dos conceptos complementarios ..	6
La pobreza de los aprendizajes.....	6
La pobreza de las evaluaciones educativas estandarizadas.....	8
Implicaciones de la pobreza de las evaluaciones educativas estandarizadas.....	10
La pobreza de las evaluaciones estandarizadas en el contexto costarricense .....	11
Superando la pobreza de las evaluaciones educativas estandarizadas .....	13
Características óptimas de una prueba estandarizada para la mejora educativa.....	19
Evidencia y análisis de un caso: Las pruebas nacionales sumativas estandarizadas de Matemática y de Español para sexto grado y para último año de secundaria en la modalidad académica diurna del año 2023 .....	21
Fortalezas y debilidades de las pruebas nacionales sumativas 2023.....	21
Análisis de la Prueba de Español sumativa ordinaria para sexto año de primaria, aplicada en el 2023 por el MEP.....	27
Aspectos generales sobre el contenido de los ítems.....	27
Análisis de los datos con el modelo de Rasch.....	34
Análisis de la Prueba de matemática sumativa ordinaria para sexto año de primaria, aplicada en el 2023 .....	44
Aspectos generales sobre el contenido de los ítems.....	44
Análisis de los datos con el modelo de Rasch.....	49
Prueba de español sumativa ordinaria para último año de secundaria, modalidad académica diurna, aplicada en el 2023 .....	55
Prueba de matemática sumativa ordinaria para último año de secundaria, modalidad académica diurna, aplicada en el 2023 .....	64
Recomendaciones generales.....	72
Referencias: .....	76

## **Descargo de responsabilidad**

Esta investigación se realizó para el *Décimo Informe Estado de la Educación (2025)*. El contenido es responsabilidad exclusiva de su autor, y las cifras pueden no coincidir con las consignadas en el capítulo respectivo, debido a revisiones posteriores. En caso de encontrarse diferencia entre ambas fuentes, prevalecen las publicadas en el Informe.

## **Introducción**

El IX Informe de EE señaló que Costa Rica experimenta la peor crisis educativa de los últimos 40 años, producto de rezagos históricos, el apagón educativo y la pandemia experimentados entre el 2018 y el 2020. (EE, 2023). Esta crisis educativa comprende tres dimensiones principales (aprendizajes, inversión y gestión educativa). Asimismo, el IX informe indicó que en materia de evaluación educativa el país no cuenta con un sistema nacional de evaluación robusto, que brinde información útil para el mejoramiento educativo, y, al contrario, en los últimos años las pruebas de macroevaluación experimentaron una gran inestabilidad y cambios, entre los últimos tres gobiernos.

Este trabajo analiza, en el contexto costarricense, lo que denominamos “la pobreza de las evaluaciones educativas estandarizadas”, concepto acuñado Ad Hoc para este estudio y que pretende ser análogo al de “pobreza de los aprendizajes” desarrollado por los especialistas del Banco Mundial y UNESCO (Azevedo et al., 2021; World Bank et al.; 2022).

El argumento principal es que ambos fenómenos, pobreza de los aprendizajes y pobreza de las evaluaciones estandarizadas, están interconectados y contribuyen al deterioro educativo en Costa Rica, especialmente en el presente escenario de Crisis Educativa y creciente desigualdad económica y social.

La noción de pobreza de las evaluaciones educativas refiere a una situación en la que los instrumentos, procesos o prácticas de evaluación en el ámbito educativo resultan insuficientes, inadecuados o limitados para medir de manera válida y pertinente el aprendizaje y desarrollo integral de los estudiantes, según sean sus propósitos (formativos, sumativos, diagnósticos, de certificación y otros). Así como la pobreza de los aprendizajes hace referencia a la incapacidad

de los sistemas educativos para garantizar que todos los estudiantes adquieran conocimientos y competencias fundamentales, la pobreza de las evaluaciones educativas señala la carencia de mecanismos de evaluación que reflejen realmente el potencial, los logros y las necesidades de los estudiantes. Es más, una evaluación pobre, en vez de proveer información útil para la toma de decisiones y la mejora educativa, puede más bien contribuir a la desinformación, y comunicar, a partir de sus resultados, mensajes erróneos sobre el estado de los aprendizajes de las personas estudiantes. En el caso particular de nuestro país, sostenemos que éste ha sido un fenómeno provocado por un conjunto particular de condiciones coyunturales y estructurales, y no está asociado directamente a la falta de capacidad técnica.

De acuerdo con lo anterior, en la primera parte del presente documento se explora este concepto circunscrito a las pruebas educativas estandarizadas. Como parte de la discusión se comparan las características de una prueba de diagnóstico o formativa, con las características de una prueba para la promoción o la certificación, dado que no son intercambiables y muchos de los procesos de construcción y validación varían dependiendo, precisamente del propósito de la evaluación. Aunado a ello y considerando especialmente el propósito de contribuir al mejoramiento educativo, se propone un conjunto de cualidades o criterios que deberían utilizarse para evaluar la calidad de las pruebas educativas estandarizadas, si se busca brindar información válida y confiable para mejorar la excelencia y equidad educativas. En este conjunto de criterios se incluyen incluyendo tanto elementos técnico-científicos, como éticos.

En la segunda parte del documento, se presenta como estudio de caso para Costa Rica el análisis realizado a las pruebas nacionales estandarizadas sumativas de español y matemática aplicadas en sexto grado y en el último año de secundaria por el Ministerio de Educación en el año 2023, que se evidencian como evaluaciones pobres si el propósito es el mejoramiento educativo, pues no brindan la información pertinente para la toma de decisiones en esa dirección, y más bien comunican, a partir de sus resultados, mensajes erróneos sobre el estado de los aprendizajes de las personas estudiantes.

A nivel metodológico, para el análisis de la validez de contenido de las pruebas se contó con el valioso apoyo de dos especialistas, el Dr. Ricardo Poveda para Matemática y la Licda. Marielos Murillo para el área de comprensión lectora, quienes valoraron los ítems de las pruebas de primaria en términos de su representatividad según los programas de estudio y los grados o niveles en que se ubican dichos aprendizajes. El modelo de Rasch se utilizó para evaluar algunos aspectos de la calidad psicométrica de las cuatro pruebas analizadas, y también fue un insumo importante para evaluar la credibilidad de los niveles de desempeño que se suponen asociados a los resultados numéricos de las pruebas.

Los resultados y los ítems analizados se trabajaron a partir de bases datos anonimizadas entregadas por el MEP al Programa Estado de la Nación por orden de la Sala Constitucional del Costa Rica el 2024 (Resolución No.2024036253)

Finalmente, se plantean en el presente escrito una serie de recomendaciones a nivel macroestructural para que la evaluación estandarizada sea una herramienta para el mejoramiento educativo, recomendaciones cuya implementación depende de un esfuerzo del país como un todo, involucrando a diversas audiencias y sectores de interés, y yendo más allá de distinciones políticas partidistas o del accionar del gobierno de turno. Asimismo, también se brindan algunas recomendaciones técnicas específicas para las pruebas, para mejorar su calidad.

## **Pobreza de los aprendizajes y pobreza de las evaluaciones: dos conceptos complementarios**

### **La pobreza de los aprendizajes**

El término “Pobreza de los aprendizajes” introducido por el Banco Mundial, la UNESCO y la UNICEF para describir la incapacidad de los niños para leer y comprender un texto sencillo a los 10 años. Destaca la brecha entre el acceso a la educación y los resultados reales del aprendizaje, particularmente en los países en vías de desarrollo. (Azevedo et al., 2021; Banco Mundial, 2019).

El concepto enfatiza que, si bien muchos niños pueden matricularse en la escuela, un número significativo no alcanza las habilidades básicas de alfabetización, que son cruciales para su desarrollo y oportunidades futuras. La pobreza de aprendizajes no se trata sólo de no aprender; refleja cuestiones más amplias, como recursos educativos inadecuados, mala calidad de la enseñanza y barreras socioeconómicas que obstaculizan el aprendizaje eficaz.

El Banco Mundial (Azevedo et al., 2021) utiliza el concepto para abogar por mejoras en los sistemas educativos, centrándose en la calidad y la equidad para reducir la pobreza de aprendizajes. Abordar esta cuestión es esencial para romper el ciclo de la pobreza y fomentar el desarrollo sostenible, ya que la educación es un motor clave del crecimiento económico y el progreso social (Banco Mundial, 2019).

Asimismo, el concepto permite resaltar la brecha crítica entre el acceso a la educación y los resultados reales del aprendizaje.

De acuerdo con Azevedo et al (2021), la pobreza de los aprendizajes obstruye el desarrollo individual y social, perpetúa el ciclo de la pobreza e impide el crecimiento económico. Se requiere de intervenciones integrales para combatir este problema.

Según los datos del Banco Mundial, publicados en el informe *The state of global learning poverty* (World Bank et al.; 2022), antes de la pandemia del COVID 19 alrededor de la mitad de los niños en países de ingresos medios y bajos presentaban esta condición; sin embargo, resultados de simulaciones, basados en datos y la evidencia más recientes, muestran que la pandemia probablemente ha causado un fuerte aumento en la pobreza de aprendizaje a nivel global, alcanzando aproximadamente un 70 % y ha exacerbado las desigualdades en la educación. La modelización de simulación para este informe establece entonces que 7 de cada 10 niños en países de ingresos bajos y medianos podrían estar sufriendo actualmente de pobreza de aprendizaje (UNESCO, 2020).

En Costa Rica, el Estado de la Educación ha estudiado este tema por medio de encuestas a muestras de docentes. En el noveno y último informe del Estado de la Educación (Murillo, Araya & Barquero, 2023) se reportó que 73% de los docentes encuestados consideraba que más de la mitad de sus estudiantes de cuarto grado se encontraba en situación de pobreza de aprendizajes durante 2022. Lo anterior dado que los docentes reportaron que la mayoría de sus estudiantes no eran capaces de leer, comprender ni escribir un texto simple, acorde a su edad. Por ejemplo, el 62% de los docentes de cuarto año indicó que la mayoría de sus estudiantes no reconocen los temas centrales de un texto ni pueden leer con fluidez, y el 70% reportó graves deficiencias en la escritura de textos. Además, el 82% de los docentes percibió una desmejora en los conocimientos de lectura y escritura en comparación con generaciones anteriores.

El documento destaca que estas carencias no solo afectan el rendimiento en lengua y literatura, sino que tienen un efecto dominó sobre otras áreas fundamentales como matemáticas y ciencias, ya que la comprensión lectora es esencial para resolver problemas en estas disciplinas. La pobreza de aprendizajes, definida como la incapacidad de un niño o niña de 10 años para leer, comprender y escribir un texto simple, se ha visto agravada por el apagón educativo entre 2018 y 2021, lo que incrementó las brechas y profundizó los problemas de calidad educativa que el sistema ya arrastraba (Murillo, Araya & Barquero, 2023).

### **La pobreza de las evaluaciones educativas estandarizadas**

En el contexto de pobreza de los aprendizajes arriba señalado, se suma otro fenómeno importante que podemos definir “la pobreza de las evaluaciones educativas estandarizadas” en alusión a la incapacidad sistémica en algunos países de medir, analizar y utilizar datos generados por dichas evaluaciones de manera efectiva para el mejoramiento educativo, tanto en términos de excelencia como de equidad. El concepto se puede analizar desde dos perspectivas. Por un lado, la incapacidad sistémica de los países para utilizar las evaluaciones con propósitos de mejoramiento, y por otro lado las insuficiencias técnicas de los instrumentos y procesos técnicos de evaluación aplicados con distintos propósitos (diagnósticos, formativos, de certificación). Existe entonces una carencia de datos confiables sobre los aprendizajes de

los estudiantes y los déficits que deben ser atendidos por medio de políticas y programas que promuevan la mejora educativa. Es más, el concepto no solo refiere a la falta de validez de las mediciones realizadas en los estudiantes para la toma de decisiones, sino que, a nuestro juicio, esta también íntimamente ligado a las carencias en términos de la incorrecta o nula utilización de la evaluación de programas como disciplina científica para valorar la efectividad de las propuestas educativas, si nuestro propósito es en efecto mejorar el sistema.

Tanto la pobreza de aprendizajes como la pobreza de las evaluaciones subrayan la importancia de medir los resultados del sistema educativo para impulsar la mejora. Mientras que la pobreza de aprendizajes se centra en los déficits educativos que enfrentan los niños, la pobreza de la evaluación educativa enfatiza las insuficiencias y obstáculos de los instrumentos utilizados para evaluar adecuadamente con propósitos de mejora y enfrentar los déficits por medio de políticas y programas con efectividad demostrada, tanto en términos de equidad como de excelencia. (Picado, Conejo & Montero, 2025). Ambos conceptos resaltan la necesidad de marcos sólidos que puedan guiar el desarrollo y evaluación de políticas y programas educativos, algo de lo que lamentablemente hemos carecido en nuestro país, tal como ha sido documentado en el último informe del Estado de la Educación (Programa Estado de la Nación, 2023).

Ahora bien, hablando de la pobreza de los aprendizajes, autores como Azevedo et al (2021) señalan entre los pilares que deben erigirse para lograr su erradicación, la necesidad de contar con “una agenda ambiciosa” en términos de evaluación para la toma de decisiones, pues se tiene clara conciencia sobre la necesidad de medir adecuadamente, con validez y confiabilidad, los resultados esperados de aprendizaje, tanto para realizar el diagnóstico en cuanto a pobreza de aprendizajes de los estudiantes, como para corroborar la efectividad de las intervenciones diseñadas para atacar el problema. Desde esta perspectiva no es posible erradicar la pobreza de los aprendizajes sino se cuenta con evaluaciones de calidad dirigidas al mejoramiento educativo.

Una propuesta en torno a la necesidad de mejorar las pruebas estandarizadas también se retoma precisamente en el capítulo antes citado del noveno informe del Estado de la Educación (Murillo, Araya & Barquero, 2023), pues una de sus recomendaciones es establecer una prueba estandarizada en tercero y sexto años para evaluar las habilidades logradas en lectura y escritura, en correspondencia con el perfil de salida del Programa de Estudio de Español.

### ***Implicaciones de la pobreza de las evaluaciones educativas estandarizadas***

La falta de evaluaciones de calidad para el mejoramiento educativo tiene consecuencias relevantes para las personas y para los países. Para las personas estudiantes porque se gradúan de los diversos programas educativos con serios vacíos y deficiencias en su formación, lo que amenaza sus posibilidades de lograr mayor calidad de vida gracias a la educación. Para los países, porque sin mecanismos de evaluación robustos y válidos no hay datos confiables y es muy difícil resulta difícil identificar qué programas tienen éxito y cuáles no y cómo corregir y mejorar las prácticas docentes en las aulas. Esta falta de conocimiento puede conducir a la continuación de intervenciones ineficaces, desperdiciando recursos valiosos de los países que podrían destinarse a iniciativas de mayor impacto.

La pobreza en la evaluación para el mejoramiento educativo puede resultar en la continuación de enfoques pedagógicos que no producen resultados de aprendizaje positivos. Si los programas educativos no se evalúan sistemáticamente, es posible que los educadores no sean conscientes de los perfiles académicos de sus estudiantes, lo que perpetúa la pobreza de aprendizajes.

Además, la ausencia de evaluaciones sólidas disminuye la posibilidad de transparencia y gestar mecanismos de rendición de cuentas entre las partes interesadas, ya que no existen métricas claras para evaluar el desempeño y los resultados.

Finalmente, la pobreza de las evaluaciones conduce al "abaratamiento" de los títulos que obtienen los estudiantes porque los obtienen mediante pruebas que no logran medir debidamente todo lo que currículo nacional dice que deben saber. Los egresados salen del sistema y los déficits que arrastran impactan a los empleadores, quienes perciben una

disminución en la calidad de los graduados (sus conocimientos y habilidades) y por los docentes de los siguientes niveles de formación, particularmente en el caso de estudiantes de secundaria que ingresan a las universidades (Araki & Kariya, 2022).

### ***La pobreza de las evaluaciones estandarizadas en el contexto costarricense***

Costa Rica ha experimentado en los últimos años, una creciente pobreza de evaluaciones educativas estandarizadas con propósitos de mejoramiento, debido a factores coyunturales y estructurales. Los coyunturales están relacionados con la eliminación de las pruebas de Bachillerato en 2019, su sustitución por las pruebas FARO y su aplicación fallida en el 2022 a causa de la controversia generada por un cuestionario de factores asociados que se aplicó sin solicitar consentimiento a los padres o encargados de los estudiantes menores de edad, y luego con el nuevo gobierno, la llegada de las llamadas Pruebas Nacionales Estandarizadas en el 2023, disputadas desde su inicio debido a su opacidad metodológica, así como procedimientos y resultados evidentemente cuestionables desde la medición y evaluación educativa como disciplina científica (Programa Estado de la Nación, 2023).

En los factores de tipo estructural, destacan dos rasgos principales de la macroevaluación educativa en Costa Rica que, a nuestro juicio, pesan más que cualquiera de las que frecuentemente se mencionan en la literatura internacional.

El primero ya ha sido señalado repetidamente, inclusive en el noveno informe del Estado de la Educación (Programa Estado de la Nación, 2023), y es el doble papel que ejerce el MEP como juez y parte, al ser al mismo tiempo responsable de brindar una educación de calidad y ser evaluador a nivel macro, de los aprendizajes de los estudiantes para propósitos de promoción, en sexto grado y último año de secundaria. Esta duplicidad del MEP en términos de ser proveedor y a la vez evaluador de la educación pública, por medio de pruebas de certificación o promoción, ha generado un serio conflicto de interés a nivel político, pues claramente, bajos resultados y baja promoción en dichas pruebas se pueden ligar directamente, y se devuelven, como boomerang, a las acciones que realiza (o deja de realizar) la institución responsable por la educación pública. Por tanto, existe una fuerte presión política, sobre todo en la actualidad, en un contexto de grandes desigualdades en oportunidades educativas, para que estas

evaluaciones no reflejen adecuadamente los contenidos que deberían valorarse según los programas de estudio, sino que se midan aprendizajes más elementales que la mayoría de los estudiantes probablemente sí dominen. Esta situación ya había sido señalada también en una evaluación de las pruebas de Bachillerato realizada en el 2006 (Esquivel Alfaro et al., 2006).

El segundo factor está asociado a la relativa inexistencia de macroevaluaciones sólidas con carácter diagnóstico y formativo que provean resultados e información útil para el mejoramiento educativo por medio del accionar docente o de políticas educativas. Aunque en el pasado, previo a las reformas de FARO, el MEP construyó pruebas de diagnóstico para sexto y noveno grados, no hay evidencia de que las mismas hayan logrado incidir de manera adecuada en procesos de mejoramiento educativo. Actualmente no se utilizan pruebas de calidad dirigidas a la evaluación diagnóstica y formativa, ni se hacen esfuerzos sistemáticos para evaluar los programas y políticas educativas.

En concordancia con lo anterior, a diferencia de otros países en vías de desarrollo, en el caso de Costa Rica la falta de recursos financieros o la falta de capacidad instalada no se perfilan, desde nuestra valoración, como las razones principales de la actual pobreza de las evaluaciones educativas estandarizadas. De hecho, nuestro país, tal como se reporta en el Noveno Informe del Estado de la Educación (Programa Estado de la Nación, 2023) fue, a finales de los años 80, modelo en evaluación educativa para otros países de América Latina. Además, existen actualmente en Costa Rica pruebas educativas estandarizadas con evidencias de alta calidad técnica y validez, tal es el caso de la Prueba de Dominio Lingüístico de la UCR, y las pruebas de admisión a la UCR y al TEC.

Asimismo, la maestría en Evaluación Educativa de la UCR, desde mediados de los años 90, ha estado formando especialistas en esta área, y existe actualmente una comisión interinstitucional de estándares de calidad para las pruebas estandarizadas educativas y laborales, conformada mayormente por evaluadores educativos, la cual, entre otros productos, ha generado una importante publicación metodológica donde se detallan los procesos a seguir en la construcción y validación de las pruebas estandarizadas, titulada “Cuaderno 9:

Estándares de calidad para pruebas estandarizadas de alto impacto en el contexto académico y profesional costarricense”. (Moreira et al., 2022).

### ***Superando la pobreza de las evaluaciones educativas estandarizadas***

La evaluación educativa, como disciplina científica, constituye el marco de referencia fundamental para conocer diversos aspectos de los aprendizajes y las situaciones del estudiantado. En el caso de la macro evaluación, las pruebas estandarizadas son esenciales para monitorear a gran escala el desarrollo de los aprendizajes, así como para diagnosticar y para medir la calidad de la oferta educativa que se brinda. Es así como las pruebas estandarizadas educativas poseen implicaciones para la vida y el desarrollo de las personas. (Moreira et al., 2022; Montero-Rojas, 2013).

Específicamente, las pruebas estandarizadas son exámenes que se construyen, aplican y califican en igualdad de condiciones y evalúan los mismos aprendizajes en todas las personas examinadas. Se deben diseñar y validar siguiendo principios fundamentales de rigor científico y ética. Además de los exámenes escritos, pueden incluir evaluaciones de desempeño o de ejecución.

Estas pruebas deben tener propósitos diferentes, y complementarios, a las construidas en el aula por la persona docente, como parte de su quehacer, y utilizadas para la gestión de los procesos de aprendizaje con sus estudiantes.

Los resultados de una prueba estandarizada, bien construida y debidamente validada, brindan una línea base para evaluar el desempeño de las personas estudiantes en relación con las metas de aprendizaje establecidas en los programas de estudio.

Las pruebas estandarizadas educativas se utilizan con propósitos formativos cuando buscan identificar fortalezas y áreas de mejoramiento en los aprendizajes, de tal forma que puedan implementarse acciones para el beneficio de la persona estudiante. La función formativa no busca fiscalizar ni castigar, sino dar luces para precisar las áreas de refuerzo y mejora. Utilizadas con propósitos formativos, las pruebas deben ofrecer a cada estudiante y sus

correspondientes docentes, una descripción de sus fortalezas y áreas de mejora en aprendizajes.

Como parte de este enfoque formativo es necesario recalcar que las causas de un bajo rendimiento educativo son múltiples, y las soluciones para estos problemas deben gestarse con la participación de todas las personas ciudadanas, así como de entidades públicas y privadas.

Las autoridades educativas y otros sectores involucrados deben utilizar los resultados de estas pruebas para crear programas y políticas educativas que busquen tanto la excelencia como la equidad en la educación. Excelencia en cuanto al logro de las metas educativas establecidas en los programas de estudio. Equidad en cuanto a mitigar las diferencias en las oportunidades educativas según variables sociodemográficas o geográficas del estudiantado.

En particular, no es pertinente, en este escenario, responsabilizar a las personas docentes por la crisis educativa que atraviesa nuestro país, pues día a día deben enfrentar condiciones que limitan e interfieren con el alcance de su labor. (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014).

Para superar la situación de pobreza de las evaluaciones estandarizadas se debe resaltar la importancia de las pruebas formativas o de diagnóstico, recurriendo primeramente al concepto moderno de validez de un instrumento de evaluación educativa, acuñado por Samuel Messick (1989):

*“La validez es un juicio evaluativo integrado del grado en que la evidencia empírica y las justificaciones teóricas apoyan la adecuación y pertinencia de las interpretaciones y acciones basadas en los puntajes de pruebas u otros tipos de medición”.*

El modelo de Messick es fundamental para la evaluación educativa porque proporciona un marco integral y ético para la validación de instrumentos de medición, asegurando que sus resultados sean interpretados y utilizados de manera adecuada y responsable en función de los

finés de la prueba (qué se busca con ella, para qué deseamos usarla), la evidencia disponible y el contexto de aplicación.

Entonces, la validez de cualquier prueba o examen está asociada al grado de credibilidad que poseen las interpretaciones o conclusiones derivadas a partir de sus resultados, según los propósitos establecidos (Moreira et al., 2022; American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014).

Por ejemplo, afirmar, a partir de un puntaje de 90% en un examen final de Matemática de sexto, que la persona estudiante posee un dominio avanzado de los objetivos de aprendizaje de este nivel, según los programas de estudio vigentes, es una conclusión que será certera en la medida que existan suficientes evidencias de la validez del examen, que apoyen la credibilidad de dicha interpretación.

Este marco de referencia claramente establece que son los propósitos de la prueba los que deben orientar todo el proceso de construcción, validación e interpretación, de ahí la importancia de distinguir entre pruebas estandarizadas de diagnóstico o formativas y pruebas estandarizadas de certificación o promoción. Un mismo instrumento no puede usarse para ambos propósitos de forma óptima, ni es correcto denominar al mismo instrumento prueba diagnóstica únicamente porque sus resultados no influyen sobre la promoción. Las pruebas de diagnóstico y de certificación utilizadas por distintos países en el mundo sirven finalidades diferentes y por tanto presentan diferencias importantes en términos de construcción, validación e interpretación.

El cuadro 1 compara las características principales de las pruebas estandarizadas educativas con propósitos de promoción o certificación frente a aquellas con fines formativos o diagnósticos.

**Cuadro 1**

**Comparación entre pruebas educativas estandarizadas de certificación o promoción versus pruebas diagnósticas o formativas**

<b>Característica</b>	<b>Pruebas de Certificación o Promoción</b>	<b>Pruebas Diagnósticas o Formativas</b>
<b>Propósito principal</b>	Determinar si el estudiante cumple con los requisitos para avanzar de nivel o graduarse.	Identificar fortalezas, debilidades y necesidades de aprendizaje del estudiante. Identificar vacíos de aprendizaje y mejorar procesos pedagógicos.
<b>Momento de aplicación</b>	Al final de un ciclo educativo (trimestre, año, etapa).	Al inicio, durante o al final de un proceso de aprendizaje.
<b>Frecuencia</b>	Generalmente una sola vez por ciclo o nivel.	Varias veces durante el proceso educativo.
<b>Tipo de retroalimentación</b>	Generalmente sumativa; indica si se aprueba o no.	Formativa; orientada a mejorar el aprendizaje y la enseñanza.
<b>Impacto en el estudiante</b>	Decisivo: puede determinar promoción, titulación o repetición.	No decisivo; busca orientar y apoyar el aprendizaje.
<b>Enfoque de los resultados</b>	Certificación de logros alcanzados.	Detección de áreas a reforzar y potencialidades.
<b>Uso de los resultados</b>	Decisión administrativa (promoción, certificación).	Planificación de estrategias pedagógicas y apoyo personalizado.
<b>Cobertura de contenidos</b>	Muestra representativa de objetivos terminales. Se mide lo mismos en todos los examinados	Selección estratégica y exhaustiva de contenidos clave. Debe brindarse información útil a todos los estudiantes en el continuum de habilidad  Uso de <i>Matrix Sampling</i> implica que no todos reciben los mismos contenidos.
<b>Cobertura de población</b>	Censales	Censales o muestrales (censales para dar retroalimentación individual a cada

Característica	Pruebas de Certificación o Promoción	Pruebas Diagnósticas o Formativas
		estudiante, muestrales objetivos de política educativa)
<b>Objetivo psicométrico de la medición</b>	Maximizar la información psicométrica en el punto de corte	Brindar información óptima a lo largo del continuum de habilidad
<b>Longitud</b>	Extensas si se pretende validez de contenido	Relativamente más cortas. La validez de contenido se logra por medio del uso de modelos de <i>Matrix Sampling</i> e imputación de puntajes
<b>Modelos de equiparación de puntajes en el tiempo</b>	No es necesaria su implementación, pues las decisiones se basan en criterios absolutos de desempeño, no se requiere comparar tendencias	Es imprescindible el uso de modelos de equiparación en el tiempo, pues se deben monitorear las tendencias de los indicadores de rendimiento eliminando posibles diferencias de dificultad entre las pruebas aplicadas en diferentes momentos
<b>Interpretación de resultados</b>	Puntajes numéricos absolutos Criterios de corte fijos	Niveles de desempeño con descripciones detalladas de los aprendizajes logrados y no logrados
<b>Rendición de cuentas</b>	De la persona examinada	Del programa académico, de las instituciones educativas, del sistema educativo

Fuente: Elaboración propia.

Frente al actual panorama educativo que presenta Costa Rica, se subraya la importancia de realizar evaluaciones diagnósticas válidas y confiables, utilizando pruebas estandarizadas. Estas evaluaciones permiten a los docentes identificar el estado real de los aprendizajes de sus estudiantes al inicio del ciclo lectivo, facilitando la toma de decisiones pedagógicas informadas y el diseño de estrategias de nivelación adaptadas a las necesidades detectadas. Las pruebas

estandarizadas, cuando son bien diseñadas y validadas, ofrecen una línea base objetiva para comparar el desempeño estudiantil con los estándares nacionales y permiten monitorear la efectividad de las políticas educativas y las intervenciones de mejora. Además, contribuyen a la equidad al identificar y visibilizar brechas asociadas a factores socioeconómicos o geográficos, permitiendo así orientar recursos y acciones hacia los grupos más vulnerables. En síntesis, la evaluación diagnóstica con pruebas estandarizadas es una herramienta clave para enfrentar la crisis de aprendizajes, ya que proporciona información precisa y comparable que sustenta la toma de decisiones a nivel de aula, centro educativo y sistema nacional.

Si bien es cierto también es posible extraer información útil para la toma de decisiones pedagógicas a partir de pruebas de certificación adecuadamente construidas, es necesario señalar que, si el propósito fundamental es el mejoramiento educativo, la utilización de pruebas diagnósticas o formativas es la estrategia más adecuada y efectiva. Esta información es crucial para que docentes y centros educativos puedan adaptar sus estrategias de enseñanza, personalizar el apoyo y diseñar intervenciones específicas que respondan a las necesidades reales de aprendizaje.

Las pruebas formativas correctamente construidas y validadas, deben ofrecer retroalimentación y fomentar procesos de mejora, tanto para los estudiantes como para los docentes, promoviendo la reflexión, la autorregulación y el desarrollo de competencias más profundas. En última instancia, el uso de pruebas diagnósticas debe contribuir a la equidad educativa, ya que permite intervenir oportunamente para evitar el rezago y potenciar el desarrollo integral de todos los estudiantes, alineándose con los objetivos de una educación inclusiva y de calidad.

De hecho, en la mayoría de los países con más avanzados sistemas de macro evaluación educativa, y ciertamente en el ámbito latinoamericano, se realizan evaluaciones diagnósticas y formativas con pruebas estandarizadas dirigidas a los estudiantes de primaria y secundaria con la finalidad de lograr evidencia para la mejora educativa o bien valorar la efectividad de las propuestas de mejora. Aunque, como se indicó anteriormente, y se muestra en el cuadro 1,

también es posible extraer información útil para la toma de decisiones pedagógicas a partir de pruebas de certificación adecuadamente construidas, si bien de forma más modesta.

### **Características óptimas de una prueba estandarizada para la mejora educativa**

A continuación, en el cuadro 2, se elementos clave a considerar para valorar la calidad de una prueba estandarizada educativa si el propósito es mejorar la educación, tanto en términos de excelencia como de equidad.

#### Cuadro 2

#### Características óptimas de una prueba estandarizada educativa cuando el propósito es la mejora del sistema educativo

Número	Característica
1	Ser construida y validada con suficientes evidencias de rigor científico y transparencia metodológica, según su propósito, y bajo el marco de referencia de la evaluación educativa como disciplina científica.
2	Estar elaborada por equipos humanos con la formación académica y experiencia adecuadas, incluyendo, por supuesto, profesionales en evaluación y medición educativa, psicometría, estadística y educación especial.
3	Carecer de manipulación política que afecte la integridad de la prueba, tanto a nivel científico como ético.
4	Tener propósitos de diagnóstico y formativos, sin consecuencias directas en la promoción del estudiantado de la educación regular en primaria y secundaria, hasta tanto el Estado implemente medidas concretas y de comprobada efectividad, para mitigar las crecientes desigualdades en oportunidades educativas.
5	Estar alineada con el currículo, es decir, los contenidos, habilidades y objetivos evaluados en la prueba deben corresponder con lo que se enseña, según los planes de estudios.
6	Cubrir adecuadamente los objetivos y contenidos de los programas de estudio (validez de contenido) por medio de diseños apropiados, como <i>matrix sampling</i> .

- 
- 7 Contar con una Tabla de Especificaciones adecuadamente diseñada: Este documento detalla los contenidos y procesos evaluados, incluyendo la ponderación y puntuación para cada ítem. Debe validarse previamente mediante un proceso riguroso e interdisciplinario y ser conocido por los estudiantes con suficiente antelación. Es importante recalcar que la tabla de especificaciones no es simplemente un temario, sino que indica qué y cómo se evaluará cada contenido o aprendizaje, así como si respectiva ponderación o importancia en la prueba.
- 
- 8 Contar con estándares educativos basados en los programas de estudio y debidamente validados, que sirvan de insumo para la construcción de la tabla de especificaciones. Estos estándares se deben divulgar con adecuada anticipación al personal docente, al estudiantado, sus familias y a los diversos *stakeholders* del sistema educativo.
- 
- 9 Incluir ítems, preguntas o ejercicios cuya calidad técnica se haya confirmado previamente, mediante aplicaciones piloto.
- 
- 10 Contar con los apoyos apropiados para hacerlas plenamente accesibles a las personas que los requieren, ya sea por aprendizaje, por situación de discapacidad o por pertenecer a grupos de población especialmente vulnerables.
- 
- 11 Implementar diseños de equiparación de puntajes (*equating*) que permitan la comparabilidad entre resultados, para dar seguimiento al rendimiento del estudiantado a lo largo del tiempo y según poblaciones diversas.
- 
- 12 Desarrollar pruebas piloto antes de la aplicación operacional o definitiva, en una muestra representativa de la población meta, lo anterior es crucial para validar los ítems y evaluar la confiabilidad y validez de la prueba.
- 
- 13 Incluir, como parte de los resultados, las descripciones detalladas, y debidamente validadas, de los aprendizajes logrados y no logrados por cada estudiante (niveles de desempeño), además de la nota numérica. Estas descripciones cualitativas del desempeño de la persona estudiante deben señalar de manera precisa qué puede o no hacer el estudiante según su puntuación. Lo anterior es especialmente útil para la toma de decisiones acerca de políticas educativas que se requieran para reforzar aprendizajes o para garantizar los aprendizajes en las poblaciones más vulnerables de estudiantes.
-

- 
- 14 Contar con un reporte de resultados a nivel individual de cada estudiante, además de los resultados globales por centro educativo, área geográfica y cualquier otra categoría de interés para la toma de decisiones.
- 
- 15 Abordar las inequidades de alfabetización tecnológica: Se debe reconocer y minimizar los problemas de validez y justicia generados por las desigualdades en alfabetización tecnológica, que provocan discriminación de facto para algunos grupos de estudiantes en las pruebas digitales. Es necesario valorar previamente el conocimiento y acceso a la tecnología de cada estudiante.
- 
- 16 Considerar las diferencias en oportunidades educativas: La interpretación de los resultados debe considerar las diferencias en oportunidades de aprendizaje de diversos grupos de estudiantes especialmente vulnerables debido a sus condiciones socioeconómicas o culturales.
- 

Fuente: Elaboración propia.

## **Evidencia y análisis de un caso: Las pruebas nacionales sumativas estandarizadas de Matemática y de Español para sexto grado y para último año de secundaria en la modalidad académica diurna del año 2023**

### **Fortalezas y debilidades de las pruebas nacionales sumativas 2023**

En el 2022 el CSE órgano encargado de la definición de las políticas educativas en Costa Rica decidió eliminar la Pruebas FARO como ya se indicó arriba, y acordó desarrollar nuevos exámenes que llamó Pruebas Nacionales Estandarizadas, las cuales tendrían una ponderación en las notas de promoción en sexto grado y en el último año de la educación secundaria (ya sea undécimo para los colegios académicos o duodécimo para los colegios técnicos). Estas pruebas fueron aplicadas con este propósito sumativo por primera vez en el 2023, con una ponderación de 30% en la nota de promoción para los grados antes mencionados, en el 2024 representaron un 40% de la nota de promoción, y, a partir del 2025 se estableció que su ponderación en las notas finales sería de 50%.

Si bien a la luz del concepto de pobreza de las evaluaciones estos exámenes pueden efectivamente clasificarse como pobres, a partir de los argumentos y la evidencia que en breve serán expuestos, es necesario reconocer también que existen valores agregados en estas pruebas, los cuales deberían rescatarse en futuros procesos evaluativos por medio de pruebas estandarizadas.

En concordancia con lo anterior, los siguientes son algunos aspectos positivos que presentan estas pruebas:

A nivel general, estos exámenes cumplen con el propósito de evaluar habilidades o destrezas, más allá del conocimiento meramente memorístico, tal como se plantea en los programas de estudio, bajo la transformación curricular “Educar para una nueva ciudadanía”. Precisamente esa característica, estando ahora el constructo evaluado más claramente asociado a la medición de habilidades y destrezas, es la que permite ubicar a los ítems en los niveles de desempeño de ERCE (UNESCO, 2021) y de PISA (OECD, 2023).

De igual forma los ítems están correctamente escritos, siguiendo las recomendaciones técnicas en cuanto a la construcción de preguntas de acogenia única.

También se reconoce la iniciativa de incluir reactivos del área de Estadística y Probabilidad en las pruebas de Matemática, cosa que nunca sucedió en las antiguas pruebas de Bachillerato, a pesar de que dicha área estaba presente en los programas de estudio.

La utilización de pruebas estandarizadas, sin desecharlas de plano, es también un aspecto positivo en este escenario, pues se reconoce que representan un valor agregado importante para el sistema educativo, si bien es necesario mejorar la calidad técnica de las existentes y construir pruebas de diagnóstico y formativas con la necesaria validez, partiendo de un propósito fundamental, el mejoramiento educativo, tanto en términos de excelencia como de equidad.

En cuanto a la utilización de la tecnología, la implementación de plataformas digitales para la administración de exámenes representa un avance significativo en la modernización tecnológica de los centros educativos, así como una oportunidad para incrementar la eficiencia en los procesos de evaluación. No obstante, es importante considerar que un porcentaje considerable de estudiantes aún debe realizar las pruebas en formato físico, mediante exámenes impresos, debido a la falta de infraestructura tecnológica y conectividad en sus instituciones.

Sin desmerecer lo anterior, el análisis de las Pruebas Nacionales Estandarizadas (PNE) sumativas, ordinarias, de sexto grado y de último año de secundaria de los colegios académicos diurnos en las asignaturas de español (comprensión lectora) y matemática, del año 2023, desarrolladas por el Ministerio de Educación Pública, revela importantes deficiencias técnicas, si se intenta utilizarlas con propósitos de diagnóstico, e incluso con fines de certificación o promoción, como es su fin pretendido. Estas deficiencias se pueden resumir como sigue:

- Carencia de validez de contenido en grado aceptable:
- Los aprendizajes evaluados en las pruebas son una fracción menor al 25% de todos los aprendizajes establecidos en los programas de estudio.
- El nivel de dificultad de los ítems es, en general, bastante más bajo que lo establecido por las pruebas internacionales de diagnóstico ERCE (Estudio Regional Comparativo y Explicativo) para sexto grado (UNESCO, 2021) y PISA, dirigida a estudiantes de 15 años (OECD, 2023).
- Ambas pruebas de Español, tanto la de sexto grado como la de último año de educación secundaria en la modalidad académica diurna, evalúan destrezas de comprensión lectora. Analizando las descripciones de desempeño de las pruebas de ERCE de tercer y sexto grados se concluye que la mayoría de los ítems de comprensión lectora en nuestra prueba de sexto podrían ubicarse, según ERCE, en los desempeños superiores o avanzados de tercer grado, excluyéndose del examen costarricense los niveles superiores de desempeño

de sexto grado de ERCE. Por su parte, en la prueba de Español de secundaria los desempeños evaluados por la mayoría de los ítems podrían clasificarse en los niveles 2 y 3 de los 6 niveles evaluados por PISA. En el caso de la prueba de Matemática de secundaria la situación es similar.

- Además, el análisis realizado por sendos expertos en Español y Matemática de las correspondientes pruebas de sexto grado, reveló que, de acuerdo con los programas de estudio actualmente vigentes en nuestro país, los aprendizajes evaluados en dichas pruebas de sexto grado, en su mayoría corresponden a años anteriores al sexto, en ambas asignaturas.

Debemos recordar que ERCE se basa en los marcos curriculares de los países de América Latina, y genera pruebas para tercer y sexto grados. PISA, aunque no parte de un currículo específico, evalúa las habilidades que una persona de 15 años debería poseer en alfabetización matemática y competencia lectora, edad que la gran mayoría de los estudiantes han superado al finalizar la secundaria.

- Problemas metodológicos e interpretación confusa:
- La metodología para la construcción y validación de las pruebas carece de transparencia, limitando la interpretabilidad de los resultados. El MEP no proveyó suficiente información técnica sobre la calidad de la prueba. Por ejemplo, no se conocieron los resultados de las aplicaciones piloto y operacionales en términos de las estadísticas de los ítems, tampoco se divulgó el nivel de confiabilidad de las pruebas, todos ellos son elementos importantes para valorar en una prueba que pretenda brindar información útil para el mejoramiento educativo.
- Un elemento que puede comprometer el grado de validez de las interpretaciones de los puntajes es que la llamada “prueba diagnóstica”, que se aplica meses antes de la correspondiente prueba sumativa de cada asignatura, está compuesta por ítems que son muy similares a los ítems que se presentarán luego en las pruebas sumativas. Esta circunstancia puede restarle validez a la medición de habilidades, por el posible efecto de la

memorización del contexto del ítem y de la estrategia de resolución. Por supuesto que la prueba de práctica y la prueba sumativa deben partir de la misma tabla de especificaciones, pero los ítems no deberían presentar una similitud tan evidente, en términos de contexto y de estrategias de resolución.

- Categorización imprecisa y equívoca para la interpretación de los resultados. La clasificación de los estudiantes en tres niveles de desempeño y los nombres tan contundentes usados para cada uno, "Básico," "Intermedio," y "Avanzado", no parece estar adecuadamente fundamentada, sobre todo considerando que la mayoría de los aprendizajes evaluados corresponden a años inferiores y no a sexto grado ni al último año de secundaria. El análisis realizado con el modelo de Rasch de los datos generados por las pruebas de español y matemática tampoco brinda evidencia de sustento para dicha categorización. Asimismo, este análisis no da sustento para justificar las descripciones sustantivas de desempeño en cada una de las tres categorías, considerando especialmente que se trata de pruebas estáticas, lineales, de 30 ítems en el caso de primaria y 35 ítems en el caso de secundaria, números bastante reducidos, en general, para lograr descripciones adecuadas de desempeños en una prueba estandarizada sumativa de final de año. Así las cosas, la interpretación de los resultados presentados puede contribuir a conclusiones erróneas. Como ilustración, la clasificación en "Avanzado" sugiere, incorrectamente, que la persona estudiante se encuentra en una posición superior en cuanto a los conocimientos, habilidades y destrezas que debería dominar en sexto grado o en el último año de secundaria, lo cual definitivamente no podría asegurarse a partir de los puntajes obtenidos en el examen. De hecho, estas ideas equivocadas pueden verse reforzadas al leer los informes de resultados elaborados por el MEP, cuando se indica, por ejemplo, en el dirigido a los docentes de primaria y correspondiente a la aplicación sumativa del 2023, lo siguiente:

*“Este informe, dirigido a los educadores de primaria, contiene los datos correspondientes a los niveles de desempeño alcanzados por el estudiantado de sexto año de la Educación General Básica en la aplicación sumativa de la Prueba Nacional Estandarizada 2023, con respecto a los **conocimientos y las habilidades del perfil de salida del ciclo correspondiente**, fundamentado en los programas de estudio vigentes, la Política Educativa y la Política Curricular.”*

(Ministerio de Educación Pública, 2024a, P.2) (La negrita es nuestra).

A partir de lo anterior y de lo que se discutió en las secciones previas, es claro que no es factible pretender que las pruebas estandarizadas que utiliza el MEP para propósitos de promoción en sexto grado y último año de secundaria puedan utilizarse como herramientas adecuadas para alimentar decisiones orientadas al mejoramiento educativo. Se trata de instrumentos que buscan otra finalidad, por lo que tampoco se puede aspirar a que posean las características de una prueba cuyo propósito principal es el mejoramiento educativo. Por lo anterior es que se necesita implementar de manera urgente un programa de pruebas diagnósticas y formativas, que cumplan con las características de calidad listadas en el cuadro 2 del presente documento.

A continuación, se presenta un análisis más detallado de cada una de las dos pruebas de primaria, de Español y Matemática, en donde se evidencian de forma más específica algunas de sus debilidades técnicas. El análisis para las dos pruebas de secundaria en las mismas asignaturas se presenta luego, aunque de manera menos detallada que el correspondiente a primaria.

## **Análisis de la Prueba de Español sumativa ordinaria para sexto año de primaria, aplicada en el 2023 por el MEP**

### **Aspectos generales sobre el contenido de los ítems**

Esta es una prueba que se propone evaluar comprensión lectora en estudiantes de sexto grado. Se aplica con propósitos sumativos hacia finales del año escolar y tiene incidencia directa en la promoción de cada estudiante. Está compuesta por 30 ítems de selección única, con tres opciones de respuesta cada uno. Algunos aspectos que resaltan, a partir de su valoración general, son los siguientes:

- El nivel de dificultad de la gran mayoría de los ítems que la componen es bajo, considerando los estándares internacionales en cuanto a competencias lectoras en sexto grado, correspondiendo, mayormente, a aprendizajes que debieran lograrse en tercer y cuarto años, según los programas de estudio actualmente vigentes en nuestro país (Murillo & Barquero, 2025).
- Aun limitándonos a la evaluación de la comprensión lectora, y excluyendo los otros tres pilares de la competencia comunicativa (escucha, escritura y comunicación oral), hay una restricción importante en los aprendizajes evaluados, pues se limita a tareas básicas, correspondientes al primer nivel de comprensión que consiste en identificar información de diferente tipo en los textos, según una revisión realizada por una especialista en lectoescritura (cuadro 3).
- Aunado a lo anterior, del análisis realizado por la experta en lectoescritura, solo hay un ítem entre los 30 que exige una interpretación inferencial para resolverlo, los demás pueden ser resueltos a partir de interpretaciones de la información explícita presente en el texto y de pistas evidentes (cuadro 3).

- La interpretación inferencial, así como los procesos de predicción y comprensión crítica, deben ser parte de la comprensión lectora en un nivel de sexto grado, según los estándares establecidos por ERCE (UNESCO, 2021), pero están ausentes en esta evaluación, aun cuando los programas de estudio sí los consideran.
- Está ausente también la comprensión lectora de imágenes, gráficos, cuadros y esquemas, habilidad fundamental del mundo moderno y que sí se encuentra presente de manera explícita en los programas de estudio.
- Hay una posibilidad de sesgo en la medición del constructo, puesto que los ítems de textos literarios se basan en extractos de obras que los niños ya conocen, dado que son parte de sus lecturas propuestas por el Ministerio de Educación. El conocimiento previo es entonces un factor que amenaza también la validez de las interpretaciones. Por lo cual la sugerencia es utilizar lecturas no conocidas previamente por los estudiantes, si lo que se intenta medir es competencia lectora.
- Debemos recordar que, en términos generales, la longitud de un texto está asociada a su dificultad, y en sexto grado debería esperarse que los estudiantes sean capaces de leer y comprender textos de al menos una página de longitud. Sin embargo, los textos de los ítems en esta prueba son bastante cortos, todos ellos tienen longitudes menores a las 140 palabras y varios incluyen apenas unas pocas líneas.

Cuadro 3

Clasificación de los ítems de la prueba sumativa ordinaria de español de sexto grado realizada por una experta en lectoescritura

Ítem	¿Qué se pide?	Tarea específica	Porcentaje de aprobación	Teórico	Año escolar del ítem según MEP (2012) (criterio de experto)
Ítem 1	Ideas fundamentales	identificación de un dato	65	Básico	3 y 4
Ítem 2	Ideas fundamentales	Síntesis de un contenido explícito	67	Intermedio	3 y 4
Ítem 3	Ideas fundamentales	identificación de una idea explícita	64	Básico	3 y 4
Ítem 4	Ideas fundamentales	Síntesis de un contenido explícito	74	Intermedio	3 y 4
Ítem 5	Idea complementaria	Deducción de una causa	54	Intermedio	No hay nivel de desempeño básico de ideas complementarias 3 y 4
Ítem 6	Idea complementaria	Forma literal sinonímica	80	Básico	3 y 4
Ítem 7	Idea complementaria	Forma literal explícita	73	Básico	3 y 4
Ítem 8	Idea complementaria	Forma literal explícita	80	Básico	3 y 4
Ítem 9	Relaciones de causa	Causa única explícita/texto expositivo/Idea complementaria	79	Básico	3 y 4
Ítem 10	Relaciones de causa	Causa única explícita/texto descriptivo	88	Básico	3 y 4

*Pobreza de las evaluaciones educativas: El caso de las pruebas estandarizadas sumativas de sexto grado y último año de secundaria en Español y Matemática aplicadas en Costa Rica en el 2023*

Ítem	¿Qué se pide?	Tarea específica	Porcentaje de aprobación	Teórico	Año escolar del ítem según MEP (2012) (criterio de experto)
Ítem 11	Relaciones de causa	Causa única explícita/texto expositivo	82	Básico	3 y 4
Ítem 12	Relaciones de causa	Causa única explícita/texto descriptivo	82	Básico	3 y 4
Ítem 13	Relaciones de efecto	Consecuencia única explícita de un texto no literario	84	Básico	MEP las puso únicamente en nivel básico 3 y 4
Ítem 14	Relaciones de efecto	Consecuencia única explícita de un texto no literario	84	Básico	MEP las puso únicamente en nivel básico 3 y 4
Ítem 15	Relaciones de efecto	Consecuencia única explícita de un texto no literario	89	Básico	MEP las puso únicamente en nivel básico 3 y 4
Ítem 16	Relaciones de efecto	Consecuencia única explícita de un texto no literario	89	Básico	MEP las puso únicamente en nivel básico 3 y 4
Ítem 17	Reconocimiento de temas en textos literarios	Tema cotidiano explícito	86	Básico	3 y 4

*Pobreza de las evaluaciones educativas: El caso de las pruebas estandarizadas sumativas de sexto grado y último año de secundaria en Español y Matemática aplicadas en Costa Rica en el 2023*

Ítem	¿Qué se pide?	Tarea específica	Porcentaje de aprobación	Teórico	Año escolar del ítem según MEP (2012) (criterio de experto)
Ítem 18	Reconocimiento de temas en textos literarios	Tema único evento implícito	51	Intermedio	4 y 5
Ítem 19	Reconocimiento de temas en textos literarios	Tema único evento implícito	31	Intermedio	4 y 5
Ítem 20	Reconocimiento de temas en textos literarios	Tema único evento implícito	80	Básico	3 y 4
Ítem 21	Inferencia de pensamientos de los personajes	Inferencia en textos literarios	77	Intermedio	Solo en intermedio (3 y 4)
Ítem 22	Inferencia de pensamientos de los personajes	Inferencia en textos literarios	70	Intermedio	4 y 5
Ítem 23	Inferencia de pensamientos de los personajes	Inferencia en textos literarios	66	Intermedio	4 y 5

*Pobreza de las evaluaciones educativas: El caso de las pruebas estandarizadas sumativas de sexto grado y último año de secundaria en Español y Matemática aplicadas en Costa Rica en el 2023*

Ítem	¿Qué se pide?	Tarea específica	Porcentaje de aprobación	Teórico	Año escolar del ítem según MEP (2012) (criterio de experto)
Ítem 24	Reconocimiento de conflictos en textos literarios	Reconocer conflictos a nivel implícito	78	Básico	3 y4
Ítem 25	Reconocimiento de conflictos en textos literarios	Reconocer conflictos a nivel explícito	82	Básico	3 y4
Ítem 26	Reconocimiento de conflictos en textos literarios	Reconocer conflictos a nivel implícito	69	Básico	3 y4
Ítem 27	Reconocimiento de conflictos en textos literarios	Reconocer conflictos a nivel implícito	67	Básico	3 y4
Ítem 28	Comportamiento de personajes en textos literarios	Identifica a nivel implícito de una situación única	85	Intermedio	4 y 5
Ítem 29	Comportamiento de personajes en textos literarios	Identifica a nivel implícito de una situación única	56	Intermedio	4 y 5

*Pobreza de las evaluaciones educativas: El caso de las pruebas estandarizadas sumativas de sexto grado y último año de secundaria en Español y Matemática aplicadas en Costa Rica en el 2023*

---

Ítem	¿Qué se pide?	Tarea específica	Porcentaje de aprobación	Teórico	Año escolar del ítem según MEP (2012) (criterio de experto)
Ítem 30	Comportamiento de personajes en textos literarios	Identifica a nivel implícito de una situación única	75	Intermedio	4 y 5

---

Fuente: Murillo y Barquero, 2025.

## **Análisis de los datos con el modelo de Rasch**

El modelo de Rasch ofrece ventajas clave tanto en la interpretación como en la aplicación práctica para el análisis de resultados en pruebas educativas. Debido a estos beneficios, su uso se ha extendido ampliamente en programas internacionales de evaluación y, de manera general, en iniciativas nacionales que emplean pruebas estandarizadas. En particular, las ventajas del modelo de Rasch pueden aplicarse en evaluaciones como las pruebas PISA, pruebas de diagnóstico y exámenes de certificación. Entre sus principales aportes se encuentran los siguientes (Montero-Rojas & Rojas-León, 2012; Bond & Fox, 2015):

- A partir de la puntuación obtenida mediante el modelo de Rasch, es posible determinar de manera sustantiva los logros del estudiante en relación con los procesos y contenidos evaluados en la prueba, así como identificar sus áreas de mejora en función del constructo medido.
- Se calcula un margen de error para cada puntuación estimada dentro del constructo de interés, lo que permite mejorar la calidad de las pruebas. Esto se debe a que se cuenta con herramientas objetivas que facilitan la adaptación de los exámenes a los propósitos específicos y a los grupos a los que están dirigidos.
- Además de evaluar el ajuste de los ítems según sus características psicométricas, el modelo de Rasch permite analizar el comportamiento de los examinados, identificando patrones que sugieran respuestas al azar, falta de atención o incluso posibles intentos de copia.
- La medición conjunta es la propiedad más importante de este modelo y posibilita representar en una misma escala tanto la dificultad de los ítems como el nivel del atributo que se busca evaluar en los examinados, ya que ambos parámetros están dados en las mismas unidades (logits). Lo anterior permite establecer una correspondencia entre el nivel del constructo reflejado en los ítems y el nivel del constructo que poseen los estudiantes. Dichos resultados se presentan en un gráfico denominado **mapa de personas versus ítems**. A partir de esta característica surge el valor diagnóstico del modelo de Rasch, pues es

posible generar, con el aporte de especialistas, los niveles de desempeño, los cuales sirven como referencia para comparar los resultados de los examinados.

En este contexto se analizaron los datos generados por la prueba a partir de su aplicación censal que contó con un total de 65535 estudiantes. Los archivos de datos ya venían codificados, asignando un cero si la respuesta a cada ítem era incorrecta y un 1 si era correcta.

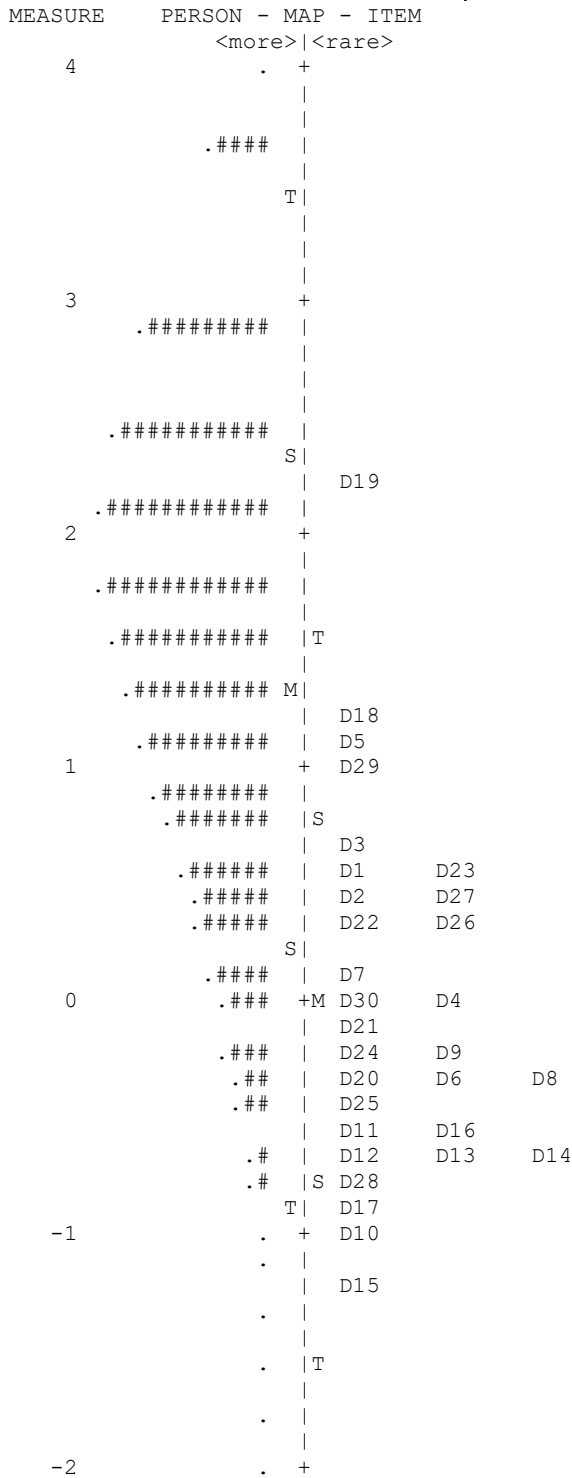
Los resultados del mapa de personas versus ítems se presentan en la figura 1.

Figura 1

Mapa de personas versus ítems según el modelo de Rasch

Prueba Nacional Estandarizada sumativa ordinaria de español de sexto grado

N= 65535, año 2024, confiabilidad de personas = 0.74



Fuente: Elaboración propia con datos suministrados por el MEP.

Los resultados del modelo de Rasch muestran un ajuste aceptable para los ítems, con valores de INFIT entre 0.84 y 1.47. Bond y Fox (2015) afirman que las magnitudes recomendables para INFIT deben estar en el intervalo [0.8 -1.2].

En este caso solamente el ítem 19 muestra algún desajuste, siendo además el más difícil del conjunto, y representando claramente un valor extremo o “outlier” en la tendencia general de los datos, según muestra el mapa de personas versus ítems. Este comportamiento podría deberse, precisamente, a que es el único ítem que requiere de una interpretación inferencial para responderse.

De la observación del mapa de personas versus ítems se concluye que la distribución de los estudiantes respondientes tiende a concentrarse hacia puntajes relativamente altos, mientras que la gran mayoría de los ítems se concentran en valores bajos de dificultad. De hecho, cerca de la mitad de los examinados poseen estimaciones del constructo que están por encima de la dificultad de los ítems, con la excepción del susodicho ítem 19 en el formulario del examen.



De hecho, posiblemente debido a esta desarticulación entre los niveles de dificultad de los ítems y las estimaciones de la habilidad de los estudiantes, es que la confiabilidad de las personas no llega al 0.80 requerido para la toma de decisiones sobre las personas examinadas, alcanzando apenas un valor de 0.74 (Muñiz, 2003).

A partir de estos datos la tarea de construir 3 intervalos de puntaje se vuelve estadísticamente problemática, pues al ser los ítems relativamente fáciles para una mayoría de la población, solamente el nivel básico parece poseer suficientes ítems para representarse de manera adecuada. Sin embargo, los niveles de desempeño para la interpretación de los puntajes establecidos por el MEP fueron tres, y definidos en la figura 2, tal como sigue:

Figura 2

Niveles de desempeño, según el MEP, para la prueba sumativa ordinaria de sexto grado de español, año 2023

## DESCRIPTORES DE LOS NIVELES DE DESEMPEÑO POR COMPONENTE DE LA PRUEBA

	 <b>Básico</b>	 <b>Intermedio</b>	 <b>Avanzado</b>
 <b>Español</b>	Distingue las ideas fundamentales en textos no literarios, a partir de la tipología textual argumentativa.	Distingue las ideas fundamentales en textos no literarios, a partir de la tipología textual informativa.	Distingue las ideas fundamentales en textos no literarios, a partir de la tipología textual expositiva y argumentativa.
	Distingue las ideas complementarias en textos no literarios, a partir de información explícita con vocabulario básico.	Determina la relación de causa en textos no literarios, a partir de la tipología textual expositiva que demanda una interconexión de ideas explícitas.	Distingue las ideas complementarias en textos no literarios, a partir de información implícita con vocabulario técnico.
	Determina la relación de causa en textos no literarios, a partir de la tipología textual expositiva con información a nivel literal.	Reconoce los conflictos de los personajes en textos literarios, a partir de varios acontecimientos en el nivel explícito.	Determina la relación de causa en textos no literarios, a partir de la tipología textual expositiva que demanda un proceso deductivo.
	Reconoce los conflictos de los personajes en textos literarios, a partir de un acontecimiento en el nivel explícito.	Identifica comportamientos de los personajes en un texto literario, a partir de información implícita.	Reconoce los conflictos de los personajes en textos literarios, a partir de varios acontecimientos en el nivel implícito.
	Identifica comportamientos de los personajes en un texto literario, a partir de contextos cotidianos.	Determina la relación de efecto en textos no literarios, a partir de información a nivel explícito.	Determina la relación de efecto en textos no literarios, a partir de la síntesis de una información.
	Reconoce los temas presentes en textos literarios, a partir de un contexto que requiere la comprensión de información explícita.	Reconoce los temas presentes en textos literarios, a partir de un contexto que requiere un proceso deductivo.	
	Infiere pensamientos de los personajes de un texto literario, a partir de la información explícita.	Infiere pensamientos de los personajes de un texto literario, a partir de la información implícita.	

Fuente: Ministerio de Educación Pública, 2024a.

Por un lado, comparando los contenidos de los ítems presentes en el cuadernillo de examen con las correspondientes descripciones de desempeño, parece que éstas no brindan la información necesaria para inferir el nivel de lectura de los estudiantes, pues realmente no es clara la progresión de la competencia lectora según se trate de niveles de desempeño 1, 2 o 3 (Básico, Intermedio y Avanzado). Asimismo, como se indicó antes, se requiere establecer características más precisas de los textos que correspondan a sexto grado para diferenciarlos de los desempeños establecidos para los años anteriores.

Por otro lado, parece claro que la gran mayoría de los ítems se podrían ubicar en el nivel superior de lectura que corresponde a tercer grado de educación primaria, bajo los estándares de ERCE. Como se recordará, ERCE elabora evaluaciones estandarizadas que se aplican a muestras representativa de estudiantes de América Latina y el Caribe, con el objetivo de medir su desempeño en diversas áreas del aprendizaje.

Estas pruebas son administradas por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), una iniciativa de la UNESCO, y localizado en Santiago de Chile (UNESCO, 2021). El estudio, a diferencia de las pruebas PISA que son “acurriculares”, se basa en lo que los países de América Latina declaran en sus marcos curriculares para los grados y áreas evaluados. Así, es posible medir los aprendizajes de los estudiantes en función de lo que los países se han propuesto lograr.

En el cuadro 4 se presentan los dos niveles de desempeño superiores (3 y 4) para tercer grado de la prueba ERCE de español aplicada en el 2019, año más reciente en que se realizó. Como puede observarse de estas descripciones, ellas encajan muy apropiadamente con las características de los ítems y las tareas que se solicita realizar a los estudiantes en la prueba de sexto grado de Costa Rica.

**Cuadro 4**

**Niveles de desempeño III y IV de la prueba ERCE de español para tercer grado, año 2019**

<b>Descripción de los niveles de desempeño en Lectura de los estudiantes de 3° grado de primaria</b>	
<p>Los estudiantes de este nivel escolar se enfrentaron en el ERCE 2019 a narraciones literarias, textos líricos, afiches, noticias, textos instruccionales, avisos y artículos informativos. Estos textos se caracterizaban por abordar temas conocidos o que son comunes para estudiantes de esta edad y desarrollarse en torno a un tema central, en su mayoría, a través de ideas explícitas. Este tema central puede apoyarse a través de claves evidentes presentes en lugares destacados del texto o en imágenes y complementarse con algunos detalles e ideas que le aportan mayor complejidad.</p> <p>La mayor parte de los textos presentaba una organización convencional respecto a su tipología, por ejemplo, cuentos que siguen la estructura prototípica de inicio, conflicto y desenlace. También se incluyeron textos discontinuos sencillos, es decir, textos que no necesitan leerse de manera lineal y que incorporan imágenes que reafirman la información escrita.</p>	
<b>Nivel IV</b> (Desde 813 puntos)	<p>Los estudiantes de este nivel mostraron evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Realizar inferencias que requieren concluir sobre el texto a partir de pistas presentes en él.</li> <li>• Realizar inferencias (ej. el tema central, las características o sentimientos de personajes, el conflicto) que requieren comprender el texto globalmente e integrar distinta información implícita presente en él.</li> <li>• Reconocer tipos de textos cuando su estructura no es prototípica.</li> <li>• Inferir el propósito de un texto cuando este no es evidente pues se debe diferenciar el propósito central de aquellos secundarios.</li> <li>• Aplicar conocimientos sobre tipologías textuales en la lectura de un texto.</li> </ul>
<b>Nivel III</b> (Entre 729 y 812 puntos)	<p>Los estudiantes de este nivel mostraron evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Localizar información o relaciones explícitas presentadas en forma literal o parafraseada (ej. a través de sinónimos) y que es necesario distinguir de otras informaciones similares.</li> <li>• Realizar inferencias que requieren seleccionar, relacionar o comparar distintas informaciones del texto.</li> <li>• Realizar inferencias (ej. inferir el tema central, las características o sentimientos de los personajes, el conflicto) que requieren comprender el texto globalmente a partir de información implícita, pero claramente sugerida en él.</li> <li>• Relacionar información visual y verbal de un texto.</li> <li>• Inferir el propósito de un texto cuando su organización lo sugiere.</li> </ul>

Fuente: UNESCO, 2021.

Analizamos ahora los 4 niveles de desempeño establecidos por ERCE para sexto grado en español. Estos se presentan en el cuadro 5.

Cuadro 5

Niveles de desempeño de la prueba ERCE de español para sexto grado, año 2019

<b>Descripción de los niveles de desempeño en Lectura de los estudiantes de 6° grado de primaria</b>	
<p>Los estudiantes de este nivel escolar se enfrentaron en el ERCE 2019 a narraciones literarias y poemas, cartas, notas, artículos informativos, noticias, relatos, afiches y comentarios. En algunos casos se les presentó un texto y en otros se presentaron dos textos juntos para fines de la evaluación, por ejemplo, para solicitarles compararlos o seleccionar alguna información. Las lecturas de la prueba de 6° grado abordaban temas que son conocidos o comunes para estudiantes de esta edad. Los textos incluidos tienden a desarrollarse en torno a un tema central que se presenta a través de ideas en su mayoría implícitas.</p> <p>La mayor parte de los textos presenta una organización convencional respecto a su tipología, sin embargo, en ocasiones se incluyen algunos que se alejan de dicha organización, por ejemplo, al combinar diferentes secuencias textuales dentro de una misma lectura (ej. una instrucción que contiene secciones explicativas y descriptivas). Al igual que en la prueba de 3° grado, en este nivel, también se incluyen textos discontinuos donde se hace uso de diferentes modos para presentar la información.</p>	
<b>Nivel IV</b> (Desde 810 puntos)	<p>Los estudiantes de este nivel mostraron evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Realizar inferencias a partir de conexiones entre ideas específicas o poco predecibles y ubicadas en diferentes partes de uno o más textos.</li> <li>• Realizar inferencias (ej. inferir el tema central, las características de personajes, el conflicto y el desenlace) que requieren comprender el texto globalmente y desafiar el conocimiento de mundo que él o la estudiante pueda tener.</li> <li>• Inferir el propósito de un texto a partir de claves implícitas que es necesario distinguir de información que compete.</li> <li>• Comparar dos textos (ej. su propósito comunicativo o su contenido) integrando o discriminando información que compete.</li> <li>• Evaluar la incorporación de un recurso considerando el propósito del texto.</li> </ul>
<b>Nivel III</b> (Entre 754 y 809 puntos)	<p>Los estudiantes de este nivel mostraron evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Realizar inferencias a partir de conexiones entre ideas específicas o secundarias y ubicadas en diferentes partes de uno o más textos.</li> <li>• Inferir el tema central de un párrafo o parte del texto y establecer una relación con el texto en su conjunto.</li> <li>• Realizar inferencias (ej. inferir el tema central, las características o sentimientos de personajes, el conflicto y el desenlace) que requieren comprender el texto globalmente e integrar ideas implícitas presentes en él.</li> <li>• Interpretar expresiones en lenguaje figurado a partir de claves implícitas en el texto o que desafían el conocimiento que tiene el estudiante del mundo que lo rodea.</li> <li>• Relacionar información visual y verbal de un texto.</li> <li>• Inferir el propósito de un texto a partir de claves implícitas sugeridas en el texto.</li> <li>• Comparar dos textos (ej. su propósito comunicativo o su contenido) e integrar información prominente del texto.</li> </ul>
<b>Nivel II</b> (Entre 612 y 753 puntos)	<p>Los estudiantes de este nivel mostraron evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Localizar información o relaciones explícitas presentadas en forma literal o parafraseada y ubicadas en el cuerpo del texto y distinguirlas de otras informaciones.</li> <li>• Realizar inferencias a partir de información claramente sugerida en el texto o de conexiones entre ideas que se presentan próximas entre sí.</li> <li>• Inferir el tema o idea central de un párrafo o parte del texto.</li> <li>• Realizar inferencias (ej. inferir el tema central, las características de personajes, el conflicto o el desenlace) que requieren comprender el texto globalmente y que se apoyan en claves evidentes o en su conocimiento de mundo.</li> <li>• Interpretar expresiones en lenguaje figurado cuando hay pistas evidentes o cuando se sustentan en el conocimiento que el estudiante tiene del mundo que lo rodea.</li> <li>• Reconocer un elemento de la estructura prototípica de un texto (ej. su final/desenlace o moraleja).</li> <li>• Reconocer tipos de textos e inferir su propósito cuando este es evidente.</li> <li>• Comparar dos textos (ej. su propósito comunicativo o su contenido) a partir de información reiterada o que se encuentra al inicio de ellos.</li> </ul>
<b>Nivel I</b> (Hasta 611 puntos)	<p>Este nivel agrupa a los estudiantes de más bajo desempeño en la prueba. La mayor parte de los estudiantes de este nivel mostró evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Localizar información explícita presentada en forma literal o parafraseada y en un lugar del texto que es de fácil acceso y distinguible de otras informaciones.</li> <li>• Realizar inferencias a partir de información que es de fácil acceso (ej. presentada al inicio del texto).</li> </ul>

Fuente: UNESCO, 2021.

De su valoración, parece claro que la gran mayoría de los ítems de la prueba de español de Costa Rica se ubican en el nivel 2 de desempeño, pues los aprendizajes representados en ellos se refieren a las tareas descritas en ese nivel.

Ahora bien, aún dentro del nivel 2, el segundo más bajo, existe un desempeño que no está siendo evaluado en la prueba de sexto año de Costa Rica, y es el siguiente:

- Comparar dos textos (ej. su propósito comunicativo o su contenido) a partir de información reiterada o que se encuentra al inicio de ellos.

En el caso del nivel 3, cuatro de los siete desempeños listados no están presentes en los ítems de la prueba de español, a saber:

- Realizar inferencias a partir de conexiones entre ideas específicas o secundarias y ubicadas en diferentes partes de uno o más textos.
- Interpretar expresiones en lenguaje figurado a partir de claves implícitas en el texto o que desafían el conocimiento que tiene el estudiante del mundo que lo rodea.
- Relacionar información visual y verbal de un texto.
- Comparar dos textos (ej. su propósito comunicativo o su contenido) integrando información prominente del texto.

Finalmente, es claro que el nivel de desempeño más alto de ERCE en sexto grado, el nivel 4, no está siendo evaluado en la prueba de Costa Rica. Este sería el nivel de desempeño donde se ubican tareas que, de manera válida, se podrían denominar como avanzadas. Estos son los desempeños del nivel 4:

- Realizar inferencias a partir de conexiones entre ideas específicas o poco predecibles y ubicadas en diferentes partes de uno o más textos.

- Realizar inferencias (ej. inferir el tema central, las características de personajes, el conflicto y el desenlace) que requieren comprender el texto globalmente y desafiar el conocimiento de mundo que él o la estudiante pueda tener.
- Inferir el propósito de un texto a partir de claves implícitas que es necesario distinguir de información que compite.
- Comparar dos textos (ej. su propósito comunicativo o su contenido) integrando o discriminando información que compite.
- Evaluar la incorporación de un recurso considerando el propósito del texto.

A partir de lo discutido hasta ahora se concluye que todos los anteriores elementos imposibilitan un diagnóstico válido y preciso del rendimiento académico y la identificación de las áreas que requieren atención prioritaria en los estudiantes de sexto grado en español. La falta de una evaluación estandarizada de calidad, con propósitos diagnósticos y formativos contribuye así a perpetuar la pobreza de los aprendizajes en esta asignatura.

Finalmente, se debe enfatizar que ERCE no solo se evalúa la capacidad de leer palabras correctamente, sino también la habilidad para interpretar y reflexionar sobre lo que se lee. En particular a partir de la descripción de los desempeños representados en el cuarto y último nivel, parece claro que varias de esas tareas están asociadas a lo que podemos llamar lectura o pensamiento crítico. Lamentablemente ninguno de estos desempeños se evalúa en la prueba estandarizada de español para sexto grado en Costa Rica.

## **Análisis de la Prueba de matemática sumativa ordinaria para sexto año de primaria, aplicada en el 2023**

### **Aspectos generales sobre el contenido de los ítems**

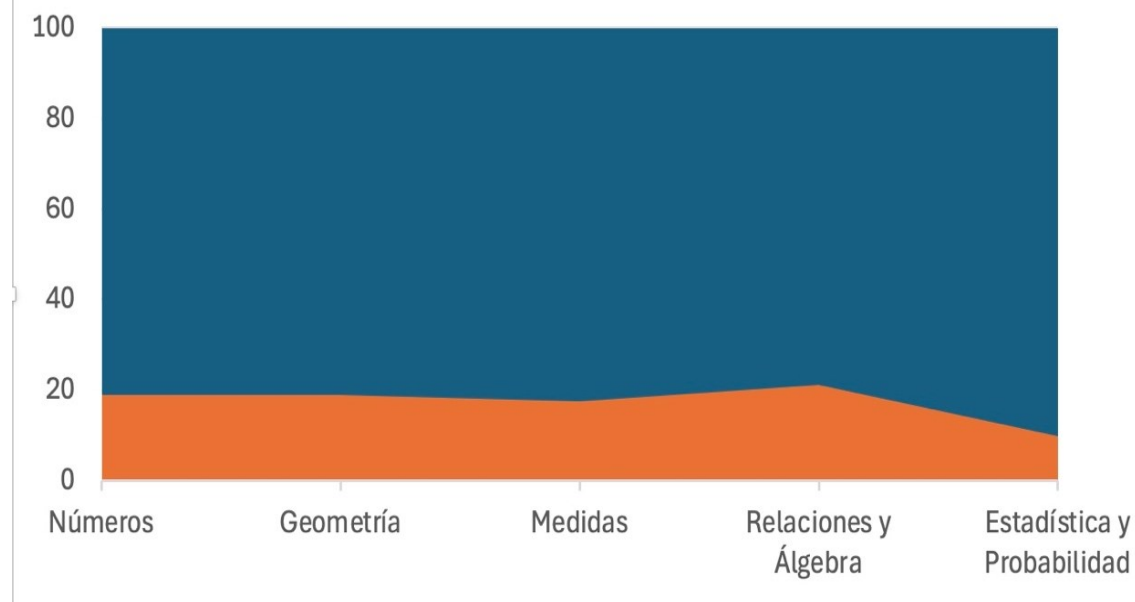
Esta es una prueba que se propone evaluar algunos conocimientos básicos de matemática que se trabajan durante la primaria. Se aplica con propósitos sumativos hacia finales del año escolar y tiene incidencia directa en la promoción de cada estudiante. Está compuesta por 30 ítems de selección única, con tres opciones de respuesta cada uno. Algunos aspectos que resaltan, a partir de su valoración general, son los siguientes:

- El gráfico<sup>1</sup> es contundente al mostrar que menos de un 25% de las habilidades de II Ciclo se evalúan en la prueba de Matemática, de hecho, tal como lo indica el análisis realizado por Zumbado et al. (2025) como especialistas en enseñanza de la matemática.
- El nivel de dificultad de la gran mayoría de los ítems que la componen es bastante bajo, considerando el análisis de la última columna del cuadro 6, donde la mayoría de los ítems se ubican en aprendizajes que corresponden a grados inferiores a sexto.
- A pesar de que los ítems tienen algunos contextos, existe una mala concepción de la resolución de problemas. Algunos ejemplos plantean contextos reales, sin embargo, las preguntas realizadas descontextualizan la situación matemática.
- Hay ítems de las cinco áreas matemáticas planteadas en el currículo nacional de matemáticas.
- A pesar de lo positivo del punto anterior, el área de Estadística y Probabilidad sigue siendo la menos considerada, solo un 10% de los ítems son de esta área.
- El promedio del porcentaje de aprobación es de un 61,5%, esto indica que a pesar de que los ítems son muy básicos y sencillos, nuestros estudiantes de primaria siguen teniendo dificultades en matemática.
- Las áreas de Estadística y Probabilidad y Relaciones y álgebra son en las que los estudiantes fallan más, por lo que es de suma importancia ponerle cuidado en la cantidad

de tiempo que los docentes están asignando en sus lecciones en estas áreas, así como el nivel de profundidad de los problemas que se trabajan en la clase.

Gráfico 1

Porcentaje de habilidades evaluadas por área matemática en la PNEM-2023 en comparación con lo planteado en MEP 2012



Fuente: Zumbado et al., 2025.

**Cuadro 6**

Clasificación de los ítems de la prueba de Matemática de sexto grado realizada por un experto en enseñanza de la matemática

Ítem	¿Qué se pide?	Tarea específica	Porcentaje de aprobación	Según MEP	Año escolar del ítem según MEP (2012) (criterio de experto)
Ítem 1	Relaciones numéricas (Números)	Leer y escribir números naturales	93,6	Básico	4
Ítem 2	Relaciones de orden (Números)	Comparar y ordenar números	19,8	Avanzado	4
Ítem 3	Potencias de base 10 (Números)	Expresar números naturales en notación desarrollada	58,1	Intermedio	6
Ítem 4	Múltiplos (Números)	Reconocer los múltiplos de un número	57,6	Intermedio	4
Ítem 5	Números compuestos (Números)	Identificar números primos y compuestos	61,7	Intermedio	6
Ítem 6	Operaciones con números decimales (Números)	Resolver problemas con sumas de números con decimales.	87,5	Básico	4
Ítem 7	Combinación de operaciones (Números)	Plantear una operación que resuelve un problema con números naturales	63,4	Intermedio	5
Ítem 8	Fracción propia e impropia (Números)	Identificar una fracción impropia	74,6	Básico	4
Ítem 9	Redondeo (Números)	Redondear un número decimal	36,5	Avanzado	5
Ítem 10	Polígonos (Geometría)	Reconocer un polígono.	76,5	Básico	4
Ítem 11	Paralelogramos (Geometría)	Reconocer propiedades de cuadriláteros referidas a los lados.	81,4	Básico	4

*Pobreza de las evaluaciones educativas: El caso de las pruebas estandarizadas sumativas de sexto grado y último año de secundaria en Español y Matemática aplicadas en Costa Rica en el 2023*

Ítem	¿Qué se pide?	Tarea específica	Porcentaje de aprobación	Según MEP	Año escolar del ítem según MEP (2012) (criterio de experto)
Ítem 12	Clasificación de triángulos (Geometría)	Reconocer un triángulo rectángulo	43,5	Intermedio	4
Ítem 13	Cilindros (Geometría)	Reconocer un cilindro	89,7	Básico	5
Ítem 14	Perímetro (Geometría)	Calcular el perímetro de un rectángulo.	66,5	Intermedio	5
Ítem 15	Área (Geometría)	Calcular el área de un cuadrado.	28,6	Avanzado	5
Ítem 16	Simetría (Geometría)	Reconocer un punto homólogo en una simetría	68,3	Intermedio	4
Ítem 17	Geometría analítica (Geometría)	Representar puntos utilizando coordenadas	64,3	Intermedio	5
Ítem 18	Tiempo (Medidas)	Medir el tiempo utilizando años, meses y días.	72,3	Básico	4
Ítem 19	Litro (Medidas)	Realizar conversiones entre el litro, sus múltiplos y submúltiplos.	42,8	Avanzado	3
Ítem 20	Moneda (Medidas)	Establecer relación entre las monedas	83,3	Básico	2
Ítem 21	Moneda (Medidas)	Establecer relación entre las monedas	60,8	Intermedio	3
Ítem 22	Sucesiones (Relaciones y álgebra)	Aplicar sucesiones y patrones para resolver problemas	69,2	Avanzado	4
Ítem 23	Sucesiones (Relaciones y álgebra)	Aplicar sucesiones y patrones para resolver problema	50,2	Avanzado	4

*Pobreza de las evaluaciones educativas: El caso de las pruebas estandarizadas sumativas de sexto grado y último año de secundaria en Español y Matemática aplicadas en Costa Rica en el 2023*

Ítem	¿Qué se pide?	Tarea específica	Porcentaje de aprobación	Según MEP	Año escolar del ítem según MEP (2012) (criterio de experto)
Ítem 24	Cantidades variables y constantes (Relaciones y álgebra)	Distinguir entre cantidades variables y constantes.	32,8	Avanzado	5
Ítem 25	Porcentajes (Relaciones y álgebra)	Resolver problemas aplicando porcentajes	35,1	Avanzado	6
Ítem 26	Regla de tres (Relaciones y álgebra)	Resolver problemas aplicando la regla de tres.	64,7	Intermedio	6
Ítem 27	Ecuaciones (Relaciones y álgebra)	Determinar el valor desconocido en una ecuación	82,4	Básico	5
Ítem 28	Muestra (Estadística y Probabilidad)	Identificar la muestra.	37,1	Avanzado	5
Ítem 29	Datos (Estadística y Probabilidad)	Identificar datos	93,1	Avanzado	3
Ítem 30	Eventos Probables (Estadística y Probabilidad)	Interpretar los conceptos de eventos menos probables.	48,6	Avanzado	3

Fuente: Zumbado et al., 2025.

### **Análisis de los datos con el modelo de Rasch**

Del examen de la figura 3, el mapa de personas versus ítems se concluye que en este caso sí existe una mayor correspondencia entre las dificultades de los ítems y las estimaciones de habilidad para los examinados, a diferencia de lo que sucede para el mapa de la asignatura de español. Sin embargo, la confiabilidad de las estimaciones de las personas es 0.73, no logrando el 0.80, valor mínimo aceptable para la toma de decisiones.

Además, resulta paradójico y llama la atención que el ítem 1 sea el más fácil y el ítem 2 el más difícil.

Los 30 ítems cumplen con el criterio de ajuste para el modelo de Rasch pues su correspondiente valor de INFIT está entre los límites aceptables (de 0.8 a 1.2).



Pasamos ahora a examinar que fundamentación brinda este análisis para los niveles de desempeño que se establecieron en esta asignatura, los cuales se presentan en la figura 4.

Figura 4

Niveles de desempeño, según el MEP, para la prueba sumativa de sexto grado de matemática, año 2024

### DESCRIPTORES DE LOS NIVELES DE DESEMPEÑO POR COMPONENTE DE LA PRUEBA




	 <b>Básico</b>	 <b>Intermedio</b>	 <b>Avanzado</b>
	<p>Interpreta las medidas de posición (moda, media aritmética, máximo o mínimo) de un grupo de datos presentados en tablas o gráficos con frecuencias absolutas, en situaciones de diversos contextos.</p> <p>Reconoce las distintas representaciones de un número natural mayor igual que 100 000, en situaciones de diversos contextos.</p> <p>Resuelve problemas de diversos contextos que involucren la suma, resta, multiplicación o división de números naturales o decimales.</p> <p>Aplica los conceptos de múltiplo de un número natural, números pares o impares, en situaciones de diversos contextos.</p> <p>Resuelve problemas de diversos contextos que involucren patrones en sucesiones con figuras o representaciones geométricas.</p> <p>Identifica el valor desconocido (letra) en una ecuación matemática dada, en situaciones de diversos contextos.</p>	<p>Reconoce si una línea corresponde o no a un polígono, en situaciones de diversos contextos.</p> <p>Aplica los conceptos de divisibilidad, divisor o factor de un número natural, en situaciones de diversos contextos.</p> <p>Determina, en diversos contextos, la conversión entre distintas medidas (longitud, peso, tiempo, capacidad, superficie, temperatura).</p> <p>Determina el término inmediato en sucesiones numéricas o tablas con números naturales menores que 1 000 000, en situaciones de diversos contextos.</p> <p>Reconoce la ubicación de puntos o figuras, en el primer cuadrante de un sistema de coordenadas, a los que se les aplicó una traslación, en situaciones de diversos contextos.</p> <p>Determina el valor desconocido (letra) en una ecuación matemática dada, en situaciones de diversos contextos.</p>	<p>Resuelve problemas de diversos contextos que involucren patrones en sucesiones numéricas o tablas con números naturales menores que 1 000 000.</p> <p>Resuelve problemas de diversos contextos que involucre el sistema monetario nacional.</p> <p>Reconoce el concepto de población o muestra, en situaciones de diversos contextos.</p> <p>Clasifica triángulos de acuerdo con las medidas de sus ángulos, en situaciones de diversos contextos.</p> <p>Reconoce la notación mixta de una fracción dada en notación impropia o viceversa, en situaciones de diversos contextos.</p> <p>Identifica trapecios o trapezoides y sus elementos (vértices, lados o ángulos), en situaciones de diversos contextos.</p> <p>Resuelve problemas de diversos contextos que involucren la conversión entre distintas medidas (longitud, peso, tiempo, capacidad, superficie, temperatura).</p>

Figura 4 (continuación)

Niveles de desempeño, según el MEP, para la prueba sumativa de sexto grado de matemática, año 2024

## DESCRIPTORES DE LOS NIVELES DE DESEMPEÑO POR COMPONENTE DE LA PRUEBA

	<b>Básico</b>	<b>Intermedio</b>	<b>Avanzado</b>
<b>Matemáticas</b>	<p>Reconoce la relación bancaria entre monedas y billetes, de todas las denominaciones del sistema monetario nacional, en situaciones de diversos contextos.</p> <p>Resuelve problemas de diversos contextos que involucren la combinación de operaciones suma, resta, multiplicación o división de números naturales.</p> <p>Calcula la medida del perímetro o área de figuras (triángulos o cuadriláteros), en situaciones de diversos contextos.</p>	<p>Identifica cubos, prismas rectangulares o cilindros, en situaciones de diversos contextos.</p> <p>Resuelve problemas que involucren el cálculo de la medida del perímetro o área de figuras (triángulos o cuadriláteros), en situaciones de diversos contextos.</p> <p>Reconoce situaciones aleatorias de diversos contextos.</p> <p>Reconoce el número decimal que corresponde a una fracción decimal, en situaciones de diversos contextos.</p>	<p>Reconoce cantidades variables o constantes, en situaciones de diversos contextos.</p> <p>Ordena números en su representación decimal, en situaciones de diversos contextos.</p> <p>Reconoce los ejes de simetría de una figura, en situaciones de diversos contextos.</p>

Fuente: Ministerio de Educación Pública, 2024a.

El planteamiento de esta tabla tan extensa y detallada de niveles de desempeño desafía la lógica de la medición educativa, pues es prácticamente imposible concluir de manera precisa acerca de desempeños tan ricos y variados, cuando solo existe un ítem como indicador de dichas tareas. En el cuadro 7 se describen en total 29 tareas diferentes, que además implican en muchos casos, su aplicación en diversos contextos o situaciones.

Sin embargo, el número total de ítems con que se cuenta para representarlas es solamente 30, y para una medición apropiada se necesitarían varios ítems para representar cada desempeño (Bond & Fox, 2015; Linacre, 2002). La tabla parece entonces estar construida a partir de lo que podríamos llamar “wishful thinking” o un argumento de fe, suponiendo que cada ítem es una

medición perfecta de los desempeños que intenta representar. Ante esta situación es claro que cualquier interpretación del rendimiento de los estudiantes basada en esta clasificación será incierta.

Además de lo anterior, y al igual que en el caso de los niveles de desempeño planteados para la prueba de español, tampoco parece clara que haya aquí una progresión lógica de complejidad o dificultad al pasar del primer al segundo nivel de desempeño y del segundo al tercero.

Es que para para lograr una tabla como la anterior que sí esté apoyada por la evidencia empírica es necesario implementar un diseño de “matrix sampling”, donde existen varios cuadernillos de examen con ítems diferentes como indicadores de las categorías de desempeño, los cuales son distribuidos aleatoriamente a los estudiantes en cada recinto donde se aplica la prueba. Por ejemplo, en las pruebas de PISA y de ERCE se usan más de 20 cuadernillos de examen con diferentes ítems para muestrear de manera adecuada los aprendizajes a evaluar y así construir luego los niveles de desempeño con la ayuda del modelo de Rasch.

Las tablas de niveles de desempeño de ERCE se construyen también a partir de la aplicación de esta metodología. Dicha tabla para la prueba de matemática de sexto grado se presenta a continuación, en el cuadro 7.

Cuadro 7

Niveles de desempeño de la prueba ERCE de matemática para sexto grado, año 2019

Descripción de los niveles de desempeño en Matemática de los estudiantes de 6° grado de primaria	
<b>Nivel IV</b> (Desde 878 puntos)	<p>Los estudiantes de este nivel mostraron evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Identificar la factorización prima de números naturales.</li> <li>• Resolver problemas que requieran calcular adiciones y sustracciones de fracciones con distinto denominador.</li> <li>• Relacionar números decimales con fracciones propias o impropias.</li> <li>• Seleccionar una ecuación de primer grado en que se utilizan símbolos en el lugar de la incógnita para modelar una situación contextualizada.</li> <li>• Clasificar cuerpos geométricos (conos, cilindros, prismas y pirámides) según sus elementos y características.</li> <li>• Resolver problemas complejos que involucren cálculo de áreas de figuras geométricas con dos o más operaciones.</li> <li>• Discriminar unidades de medida de uso poco frecuente (por ejemplo, hectáreas, decímetros cúbicos, milímetros cuadrados, etc.) que son apropiadas para medir una magnitud (longitud, masa, superficie, volumen).</li> </ul>
<b>Nivel III</b> (Entre 789 y 877 puntos)	<p>Los estudiantes de este nivel mostraron evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Resolver problemas más complejos que requieren interpretar información e involucran dos o más operaciones que incluyen multiplicación o división.</li> <li>• Interpretar el significado de variaciones proporcionales en situaciones contextualizadas.</li> <li>• Identificar fracciones equivalentes (con denominador distinto de 10) y calcular adiciones y sustracciones de fracciones con el mismo denominador.</li> <li>• Relacionar números decimales con fracciones propias sencillas o números mixtos sencillos (ej. con denominador 2) y calcular o estimar adiciones y sustracciones de números decimales.</li> <li>• Determinar términos intermedios faltantes de una secuencia presentada en una situación contextualizada, que pueden interpretar su patrón de formación.</li> <li>• Identificar relaciones de perpendicularidad y paralelismo en el plano.</li> <li>• Resolver problemas complejos que involucren cálculo o estimación de áreas y perímetros de figuras geométricas.</li> <li>• Resolver problemas que involucren medidas (masa volumen y medidas de tiempo) y convertir unidades de medidas.</li> <li>• Resolver problemas que requieren leer e interpretar información de tablas y gráficos o identificar gráficos que representen información entregada en distintos formatos.</li> </ul>
<b>Nivel II</b> (Entre 687 y 788 puntos)	<p>Los estudiantes de este nivel mostraron evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Resolver problemas simples con números naturales que involucren estimaciones o cálculos (multiplicación o división).</li> <li>• Resolver problemas más complejos (ej. que involucren una multiplicación o división) relacionados con situaciones de proporcionalidad directa.</li> <li>• Identificar representaciones gráficas de fracciones y/o fracciones equivalentes (con denominador 10).</li> <li>• Completar secuencias gráficas o numéricas complejas (ej. multiplicación) o identificar reglas o patrones de formación.</li> <li>• Resolver ecuaciones sencillas que utilicen símbolos en lugar de incógnitas.</li> <li>• Relacionar objetos del entorno con polígonos o cuerpos geométricos.</li> <li>• Resolver problemas que requieran utilizar características de cuerpos geométricos (ej. caras) para proponer soluciones de acuerdo al contexto.</li> <li>• Calcular perímetros de polígonos regulares e irregulares.</li> <li>• Organizar información en tablas o gráficos con escala.</li> </ul>
<b>Nivel I</b> (Hasta 686 puntos)	<p>Este nivel agrupa a los estudiantes de más bajo desempeño en la prueba. La mayor parte de los estudiantes de este nivel mostró evidencia de ser capaces de:</p> <ul style="list-style-type: none"> <li>• Completar secuencias numéricas simples (ej. adición) o inferir la característica común a los elementos que la componen.</li> <li>• Identificar cuerpos geométricos redondos (cono, cilindro) en objetos del entorno.</li> <li>• Relacionar una representación en perspectiva con sus posiciones relativas en un plano o mapa.</li> <li>• Estimar magnitudes (por ejemplo, longitudes) de objetos en situaciones del entorno utilizando medidas convencionales.</li> <li>• Leer datos presentados en tablas o gráficos con escala.</li> </ul>

Fuente: UNESCO, 2021.

## **Prueba de español sumativa ordinaria para último año de secundaria, modalidad académica diurna, aplicada en el 2023**

El análisis comparativo entre la prueba costarricense y la prueba de PISA revela diferencias significativas en la medición de la competencia lectora, siendo la prueba estandarizada de Costa Rica significativamente más fácil: muchas de las respuestas a los ítems están basadas en relaciones literales o sinonimias. En contraste, PISA exige en sus niveles intermedios integración de información de múltiples fuentes, elaboración de inferencias complejas y manejo de ideas abstractas y evaluación crítica con apoyo en conocimiento externo.

Mientras la prueba nacional utiliza textos continuos convencionales (fragmentos literarios y periodísticos), PISA incorpora textos mixtos con información no lineal (gráficos, tablas), textos discontinuos (formularios, infografías) materiales con contexto inusual y vocabulario especializado, géneros que exigen reconciliar perspectivas contradictorias y evaluación de la credibilidad de fuentes.

La hipótesis de que los 35 ítems de la prueba de comprensión lectora se ubican en los niveles 2 y 3 de las categorías de desempeño de PISA que se muestran en el cuadro 8, está respaldada por las características de los ítems y los criterios establecidos por el marco de evaluación.

Ningún ítem exige habilidades del nivel 4 o superior (e.g., evaluar críticamente textos técnicos o resolver problemas con información altamente conflictiva).

PISA selecciona textos que representan diferentes situaciones de la vida real, agrupadas en los ámbitos como Personal, Educativo, Público y Ocupacional. El objetivo es evaluar la capacidad de los estudiantes para comprender, emplear, valorar y reflexionar sobre textos en contextos variados y con propósitos diversos, como informarse, aprender, resolver problemas o tomar decisiones. Ejemplos concretos de textos utilizados en PISA incluyen: Artículos científicos y de divulgación (por ejemplo, sobre teorías de la Isla de Pascua) Artículos de opinión (por ejemplo, sobre el consumo de leche de vaca) Foros de debate (por ejemplo, sobre la salud de las aves), hojas informativas laborales, diagramas jerárquicos con notas al pie y otros.

Por el contrario, la prueba costarricense evalúa competencias intermedias de lectura, priorizando la interpretación de textos continuos de temas sociales y literarios con demandas cognitivas acotadas. Esto coincide con los descriptores de los niveles 2 (básico) y 3 (intermedio) de PISA, donde los estudiantes deben demostrar comprensión detallada y aplicar estrategias de análisis sin requerir evaluación crítica profunda.

Estas diferencias pueden explicar por qué Costa Rica muestra estancamiento en competencia lectora según PISA, mientras sus pruebas nacionales podrían dar falsa seguridad al medir solo habilidades básicas. La necesidad de alinear los instrumentos locales con estándares internacionales que midan pensamiento crítico y resolución de problemas complejos resulta evidente.

Los resultados del análisis con el modelo de Rasch presentados en la figura 5 brindan conclusiones similares a las que se obtuvieron para la prueba de primaria en español. De igual forma, los niveles de desempeño de la prueba costarricense listados en la figura 6 no están respaldados técnicamente, por las mismas razones que fueron argumentadas en el caso de las pruebas de primaria.

#### Cuadro 8

##### Descripción de los niveles de competencia en lectura de PISA: 2022

---

Nivel de Competencia y Puntaje de Corte Inferior	Descripción de la Tarea
<b>Nivel 6 (698)</b>	Los lectores en el Nivel 6 pueden comprender textos extensos y abstractos en los que la información de interés está profundamente integrada y solo indirectamente relacionada con la tarea. Pueden comparar, contrastar e integrar información que representa múltiples perspectivas potencialmente conflictivas, utilizando múltiples criterios y generando inferencias a través de piezas distantes de información para determinar cómo se puede utilizar la información. Los lectores en el Nivel 6 pueden reflexionar profundamente sobre la fuente del texto en relación con su contenido, utilizando criterios externos al texto. Pueden comparar y contrastar

---

**Nivel de Competencia y Puntaje de Corte Inferior**

**Descripción de la Tarea**

---

información entre textos, identificando y resolviendo discrepancias y conflictos intertextuales a través de inferencias sobre las fuentes de información, sus intereses explícitos o adquiridos, y otras pistas en cuanto a la validez de la información. Las tareas en el Nivel 6 suelen requerir que el lector establezca planes elaborados, combinando múltiples criterios y generando inferencias para relacionar la tarea con el/los texto(s). Los materiales en este nivel incluyen uno o varios texto(s) complejos y abstractos, que involucran múltiples perspectivas posiblemente discrepantes. La información objetivo puede tomar la forma de detalles que están profundamente integrados dentro o a través de los textos y potencialmente oscurecidos por información competidora.

---

**Nivel 5 (626)**

Los lectores en el Nivel 5 pueden comprender textos extensos, infiriendo qué información en el texto es relevante, aunque la información de interés pueda pasarse por alto fácilmente. Pueden realizar razonamientos causales u otras formas de razonamiento basados en una comprensión profunda de fragmentos extensos de texto. También pueden responder preguntas indirectas infiriendo la relación entre la pregunta y una o varias piezas de información distribuidas dentro o a través de múltiples textos y fuentes. Las tareas reflexivas requieren la producción o evaluación crítica de hipótesis, basándose en información específica. Los lectores pueden establecer distinciones entre contenido y propósito, y entre hecho y opinión tal como se aplica a declaraciones complejas o abstractas. Pueden evaluar la neutralidad y el sesgo basándose en pistas explícitas o implícitas pertenecientes tanto al contenido como a la fuente de información. También pueden sacar conclusiones con respecto a la fiabilidad de las afirmaciones o conclusiones ofrecidas en un fragmento de texto.

---

Nivel de Competencia y Puntaje de Corte Inferior	Descripción de la Tarea
<b>Nivel 4 (553)</b>	<p>En el Nivel 4, los lectores pueden comprender pasajes extensos en entornos de texto único o múltiple. Interpretan el significado de los matices del lenguaje en una sección del texto teniendo en cuenta el texto en su totalidad. En otras tareas interpretativas, los estudiantes demuestran comprensión y aplicación de categorías ad hoc. Pueden comparar perspectivas y extraer inferencias basadas en múltiples fuentes. Los lectores pueden buscar, localizar e integrar varias piezas de información integrada en presencia de distractores plausibles. Pueden generar inferencias basadas en la declaración de la tarea para evaluar la relevancia de la información objetivo. Pueden manejar tareas que requieren que memoricen el contexto de la tarea previa. Además, los estudiantes en este nivel pueden evaluar la relación entre afirmaciones específicas y la postura general o la conclusión de una persona sobre un tema. Pueden reflexionar sobre las estrategias que utilizan los autores para transmitir sus puntos de vista, basándose en características destacadas de los textos (por ejemplo, títulos e ilustraciones). Pueden comparar y contrastar afirmaciones hechas explícitamente en varios textos y evaluar la fiabilidad de una fuente basándose en criterios destacados. Los textos en el Nivel 4 suelen ser largos o complejos, y su contenido o forma pueden no ser estándar. Muchas de las tareas se sitúan en entornos de texto múltiple. Los textos y las tareas contienen pistas indirectas o implícitas.</p>
<b>Nivel 3 (480)</b>	<p>Los lectores en el Nivel 3 pueden representar el significado literal de textos únicos o múltiples en ausencia de contenido explícito o pistas organizativas. Los lectores pueden integrar contenido y generar inferencias tanto básicas como más avanzadas. También pueden integrar varias partes de un fragmento de texto para identificar la idea principal, comprender una</p>

Nivel de Competencia y Puntaje de Corte Inferior	Descripción de la Tarea
Nivel 2 (407)	<p>relación o interpretar el significado de una palabra o frase cuando la información requerida aparece en una sola página. Pueden buscar información basándose en indicaciones indirectas, y localizar información objetivo que no esté en una posición prominente y/o esté en presencia de distractores. En algunos casos, los lectores en este nivel reconocen la relación entre varias piezas de información basadas en múltiples criterios. Los lectores de Nivel 3 pueden reflexionar sobre un fragmento de texto o un pequeño conjunto de textos, y comparar y contrastar los puntos de vista de varios autores basándose en información explícita. Las tareas reflexivas en este nivel pueden requerir que el lector realice comparaciones, genere explicaciones o evalúe una característica del texto. Algunas tareas reflexivas requieren que los lectores demuestren una comprensión detallada de un fragmento de texto que trata sobre un tema familiar, mientras que otras requieren una comprensión básica de contenido menos familiar. Las tareas en el Nivel 3 requieren que el lector tenga en cuenta muchas características al comparar.</p> <p>Los lectores en el Nivel 2 pueden identificar la idea principal en un fragmento de texto de longitud moderada. Pueden comprender relaciones o interpretar el significado dentro de una parte limitada del texto cuando la información no es prominente produciendo inferencias básicas, y/o cuando el/los texto(s) incluyen alguna información que distrae. Pueden seleccionar y acceder a una página en un conjunto basándose en indicaciones explícitas, aunque a veces complejas, y localizar una o más piezas de información basándose en múltiples criterios, parcialmente implícitos. Los lectores en el Nivel 2 pueden, cuando se les indica explícitamente, reflexionar sobre el propósito general, o sobre el propósito de detalles</p>

Nivel de Competencia y Puntaje de Corte Inferior	Descripción de la Tarea
<b>Nivel 1a (335)</b>	<p>específicos, en textos de longitud moderada. Pueden reflexionar sobre características visuales o tipográficas simples. Pueden comparar afirmaciones y evaluar las razones que las respaldan basándose en afirmaciones cortas y explícitas. Las tareas en el Nivel 2 pueden involucrar comparaciones o contrastes basados en una sola característica en el texto. Las tareas reflexivas típicas en este nivel requieren que los lectores hagan una comparación o varias conexiones entre el texto y el conocimiento externo basándose en la experiencia personal y las actitudes.</p> <hr/> <p>Los lectores en el Nivel 1a pueden comprender el significado literal de oraciones o pasajes cortos. Los lectores en este nivel también pueden reconocer el tema principal o el propósito del autor en un fragmento de texto sobre un tema familiar, y hacer una conexión simple entre varias piezas adyacentes de información, o entre la información dada y su propio conocimiento previo. Pueden seleccionar una página relevante de un pequeño conjunto basándose en indicaciones simples, y localizar una o más piezas independientes de información dentro de textos cortos. Los lectores de Nivel 1a pueden reflexionar sobre el propósito general y sobre la importancia relativa de la información (por ejemplo, la idea principal frente al detalle no esencial) en textos simples que contienen pistas explícitas. La mayoría de las tareas en este nivel contienen pistas explícitas con respecto a lo que se debe hacer, cómo hacerlo y dónde en el/los texto(s) los lectores deben centrar su atención.</p>
<b>Nivel 1b (262)</b>	<p>Los lectores en el Nivel 1b pueden evaluar el significado literal de oraciones simples. También pueden interpretar el significado literal de los textos haciendo conexiones simples entre piezas adyacentes de información en la</p>

---

Nivel de Competencia y Puntaje de Corte Inferior	Descripción de la Tarea
Nivel 1c (189)	<p data-bbox="578 342 1458 884">pregunta y/o el texto. Los lectores en este nivel pueden escanear y localizar una sola pieza de información explícitamente establecida y colocada prominentemente en una sola oración, un texto corto o una lista simple. Pueden acceder a una página relevante de un pequeño conjunto basándose en indicaciones simples cuando hay pistas explícitas presentes. Las tareas en el Nivel 1b dirigen explícitamente a los lectores a considerar factores relevantes en la tarea y en el texto. Los textos en este nivel son cortos y típicamente brindan apoyo al lector, como a través de la repetición de información, imágenes o símbolos familiares. Hay una mínima información que compita.</p> <hr/> <p data-bbox="578 953 1458 1157">Los lectores en el Nivel 1c pueden comprender y afirmar el significado de oraciones cortas y sintácticamente simples a nivel literal, y leer con un propósito claro y simple dentro de un tiempo limitado. Las tareas en este nivel involucran vocabulario y estructuras sintácticas simples.</p>

---

NOTA: Para alcanzar un nivel de competencia particular, un estudiante debe responder correctamente a la mayoría de los elementos en ese nivel. Los estudiantes fueron clasificados en niveles de competencia en lectura de acuerdo con sus puntajes. Los puntajes de corte en la exhibición son redondeados; los puntajes de corte exactos se proporcionan en la tabla A-2 en la Metodología y las Notas Técnicas. Los puntajes se informan en una escala de 0 a 1,000.

Fuente: Organización para la Cooperación y el Desarrollo Económicos (OCDE), Programa para la Evaluación Internacional de Estudiantes. PISA, 2022.



Figura 6

Niveles de desempeño, según el MEP, para la prueba sumativa ordinaria de español, de último año secundaria, modalidad académica diurna, año 2023

## DESCRIPTORES DE LOS NIVELES DE DESEMPEÑO POR COMPONENTE DE LA PRUEBA

	 <b>Básico</b>	 <b>Intermedio</b>	 <b>Avanzado</b>
 <b>Español</b>	Identifica información en textos no literarios a partir de la reorganización de la comprensión literal, en un contexto con temática cotidiana.	Identifica información en textos literarios a partir de la reorganización de la comprensión literal, en un contexto narrativo más implícito.	Identifica información en textos no literarios a partir de la reorganización de la comprensión literal, en un contexto con temática no cotidiana y vocabulario técnico y estadístico.
	Reconoce, a un nivel explícito, el tipo de sociedad representada por el texto literario.	Identifica información en textos no literarios a partir de la reorganización de la comprensión literal, en un contexto con temática no cotidiana.	Infiere el componente semántico (implícito) en un texto no literario, a partir de un contexto que demanda información a nivel más sencillo.
	Identifica información en textos literarios a partir de la reorganización de la comprensión literal, en un contexto narrativo explícito.	Infiere el componente semántico (implícito) en un texto no literario, a partir de un contexto que demanda información a nivel más complejo.	Infiere la posición del texto literario ante el mundo mostrado, a partir de un contexto narrativo extenso.
	Infiere la posición del texto literario ante el mundo mostrado, a partir de un contexto con temática cotidiana.	Interpreta, en textos literarios, los recursos retóricos como aportes a la significación intratextual, a partir de un texto poético.	Infiere las implicaciones del pensamiento social derivadas del texto literario, a partir de un contexto de tipo argumentativo.
	Interpreta la posición del texto no literario ante el mundo representado, en un contexto con temática tecnológica.	Infiere la posición del texto literario ante el mundo mostrado, a partir de un contexto narrativo no extenso.	Interpreta, en textos literarios, los recursos retóricos como aportes a la significación intratextual, a partir de un texto narrativo.
	Reconoce el tipo de sociedad presente en el texto no literario, a partir de una inferencia y estructura sintáctica no complejas.	Reconoce el tipo de sociedad presente en el texto no literario, a partir de una inferencia y estructura sintáctica más complejas.	
Interpreta, en textos literarios, los recursos retóricos como aportes a la significación intratextual, a partir de un contexto con temática cotidiana.	Interpreta la posición del texto no literario ante el mundo representado, en un contexto con temática no tecnológica.		

Figura 6 (continuación)

### DESCRIPTORES DE LOS NIVELES DE DESEMPEÑO POR COMPONENTE DE LA PRUEBA



Fuente: Ministerio de Educación Pública, 2024b.

### Prueba de matemática sumativa ordinaria para último año de secundaria, modalidad académica diurna, aplicada en el 2023

Las características de la prueba nacional de matemática incluyen un predominio de preguntas que exigen recuperación directa de información y aplicación de fórmulas conocidas, y poca o nula presencia de problemas que requieran razonamiento matemático complejo, modelación o interpretación crítica. PISA, por su parte, evalúa la alfabetización matemática en estudiantes de 15 años con un enfoque más amplio y complejo, donde los niveles superiores implican razonamiento matemático complejo, resolución de problemas no rutinarios, uso de múltiples representaciones y conexiones entre conceptos y, pensamiento crítico y argumentación matemática.

La prueba nacional tiende a evaluar habilidades de los niveles 1, 2 y 3 de los 6 niveles evaluados por PISA, que corresponden a tareas que requieren reconocer información matemática explícita, realizar cálculos simples y aplicar procedimientos más directos. De hecho, los resultados PISA 2018 y 2022 muestran que Costa Rica tiene un puntaje promedio en matemáticas alrededor de 400 puntos, por debajo del promedio OCDE (490-500), con un alto porcentaje de estudiantes en niveles 1 y 2.

De igual forma la prueba nacional plantea problemas en contextos escolares o cotidianos, sin mucha complejidad conceptual ni interdisciplinariedad. PISA por su parte propone problemas en contextos variados (social, científico, laboral), con textos y datos no estructurados, que requieren interpretación y toma de decisiones.

La prueba nacional mide una alfabetización matemática significativamente más baja que la requerida en PISA, lo que puede explicar la brecha en resultados internacionales y la necesidad de fortalecer la evaluación y enseñanza para alcanzar estándares globales.

Los resultados del análisis con el modelo de Rasch presentados en la figura 7 generan conclusiones similares a las que se obtuvieron para la prueba de primaria en matemática. De igual forma, los niveles de desempeño listados en la figura 8 no están respaldados técnicamente, por las mismas razones que fueron argumentadas en el caso de las pruebas de primaria. Por su parte el cuadro 9 muestra los niveles de desempeño de la prueba de alfabetización matemática de PISA.



Figura 8

Niveles de desempeño, según el MEP, para la prueba de matemática sumativa ordinaria de último año de secundaria en la modalidad académica diurna, año 2023

## DESCRIPTORES DE LOS NIVELES DE DESEMPEÑO POR COMPONENTE DE LA PRUEBA

	 <b>Básico</b>	 <b>Intermedio</b>	 <b>Avanzado</b>
 <b>Matemáticas</b>	<p>Identifica elementos (monotonía) de una función, dada gráfica o algebraicamente, en situaciones de diversos contextos.</p> <p>Identifica la información que proporcionan algunas medidas de posición (el máximo y mínimo) de un grupo de datos, en situaciones de diversos contextos.</p> <p>Estima áreas de figuras planas no poligonales utilizando un sistema de coordenadas rectangulares, en situaciones de diversos contextos.</p> <p>Identifica gráficamente si un punto se ubica en el interior o en el exterior de una circunferencia, en diversos contextos</p> <p>Identifica elementos (máximo de la función (vértice), eje de simetría de la gráfica) de una función cuadrática, dada gráficamente, en situaciones de diversos contextos</p> <p>Representa algebraicamente un sistema de dos ecuaciones lineales con dos incógnitas que permita la resolución de un problema, a partir de elementos dados, de diversos contextos.</p>	<p>Interpreta la información que proporcionan algunas medidas de posición (moda, mínimo) de un grupo de datos, en situaciones de diversos contextos.</p> <p>Determina características o elementos de una función exponencial, dada algebraicamente, en situaciones de diversos contextos.</p> <p>Resuelve problemas, de diversos contextos, que involucren algebraicamente la inversa de una función lineal, cuadrática, en situaciones de diversos contextos.</p> <p>Determina la representación gráfica de una función lineal, en situaciones de diversos contextos.</p> <p>Determina la ecuación de una recta, en situaciones de diversos contextos.</p> <p>Determina características o elementos de una función logarítmica, dada algebraicamente, en situaciones de diversos contextos.</p> <p>Determina la representación algebraica de una circunferencia dado su centro y su radio, en situaciones de diversos contextos.</p>	<p>Resuelve problemas de diversos contextos donde involucren la ubicación aproximada de las medidas de posición de acuerdo con el tipo de asimetría de la distribución de los datos.</p> <p>Determina intervalos en los cuales una función representada gráficamente tiene inversa, en situaciones de diversos contextos.</p> <p>Resuelve problemas de diversos contextos donde involucren elementos (máximo) de una función cuadrática, dada gráfica o algebraicamente, en situaciones de diversos contextos.</p> <p>Relaciona la gráfica de una función con la gráfica de su inversa, en situaciones de diversos contextos.</p> <p>Resuelve problemas, de diversos contextos, que involucren el cálculo de la medida de áreas en polígonos (regulares).</p> <p>Determina elementos (preimagen) de una función, dada gráficamente, en situaciones de diversos contextos.</p>

Figura 8 (continuación)

## DESCRIPTORES DE LOS NIVELES DE DESEMPEÑO POR COMPONENTE DE LA PRUEBA

	 <b>Básico</b>	 <b>Intermedio</b>	 <b>Avanzado</b>
 <b>Matemáticas</b>	<p>Reconoce la representación gráfica de una circunferencia dado su centro y su radio, en situaciones de diversos contextos.</p> <p>Determina elementos (imagen) de una función, dada algebraicamente, en situaciones de diversos contextos.</p> <p>Reconoce la representación gráfica de la traslación de una circunferencia, en situaciones de diversos contextos.</p> <p>Determina elementos (preimagen) de la función con criterio dado por <math>f(x) = a\sqrt{x+b} + c</math>, en situaciones de diversos contextos.</p> <p>Determina características o elementos de una función logarítmica, dada algebraicamente, en situaciones de diversos contextos.</p> <p>Reconoce qué figuras se obtienen mediante secciones planas de una esfera o un cilindro, en situaciones de diversos contextos.</p>	<p>Determina geoméricamente la posición relativa entre rectas en el plano desde el punto de vista del paralelismo o la perpendicularidad, en situaciones de diversos contextos.</p> <p>Determina la medida de perímetros de polígonos, en situaciones de diversos contextos.</p> <p>Determina el tipo de función, representada tabularmente, que sirva de modelo para una situación de diversos contextos.</p> <p>Calcula áreas de polígonos no regulares utilizando un sistema de coordenadas rectangulares, en situaciones de diversos contextos.</p>	<p>Determina si un sistema de dos ecuaciones lineales con dos incógnitas tiene infinitas soluciones, en situaciones de diversos contextos.</p> <p>Resuelve problemas de diversos contextos donde involucren características métricas de secciones planas en cilindros o esferas, en situaciones de diversos contextos.</p>

Fuente: Ministerio de Educación Pública, 2024b.

**Cuadro 9**

**Descripción de los niveles de desempeño en alfabetización matemática de PISA: 2022**

**Nivel de competencia  
y puntuación de corte  
inferior**

**Descripción de la tarea**

---

**Nivel 6 (669)**

En el Nivel 6, los alumnos pueden trabajar con problemas abstractos y demostrar creatividad y pensamiento flexible para desarrollar soluciones. Por ejemplo, pueden reconocer cuándo un procedimiento no especificado en una tarea puede aplicarse en un contexto no estándar o cuándo es necesario demostrar una comprensión más profunda de un concepto matemático como parte de una justificación. Pueden relacionar diferentes fuentes de información y representaciones, incluyendo el uso efectivo de simulaciones u hojas de cálculo como parte de su solución. Los alumnos de este nivel son capaces de tener un pensamiento crítico y dominan las operaciones y relaciones matemáticas simbólicas y formales que utilizan para comunicar con claridad sus razonamientos. Pueden reflexionar sobre la idoneidad de sus acciones con respecto a su solución y a la situación original.

---

**Nivel 5 (607)**

En el Nivel 5, los alumnos pueden desarrollar y trabajar con modelos para situaciones complejas, identificar o imponer restricciones y especificar supuestos. Pueden aplicar estrategias de resolución de problemas sistemáticas y bien planificadas para abordar tareas más desafiantes, como decidir cómo desarrollar un experimento, diseñar un procedimiento óptimo o trabajar con visualizaciones más complejas que no se dan en la tarea. Los alumnos demuestran una mayor capacidad para resolver problemas cuyas soluciones a menudo requieren incorporar conocimientos matemáticos que no se indican explícitamente en la tarea. Los alumnos de este nivel reflexionan sobre su trabajo y consideran los resultados matemáticos con respecto al contexto del mundo real.

---

**Nivel 4 (545)**

En el Nivel 4, los alumnos pueden trabajar eficazmente con modelos explícitos para problemas concretos complejos. Los alumnos de este nivel comienzan a trabajar con aspectos del pensamiento crítico, como la evaluación de la

**Nivel de competencia y puntuación de corte inferior**

**Descripción de la tarea**

razonabilidad de los resultados mediante juicios cualitativos cuando no es posible realizar cálculos a partir de la información proporcionada. Los alumnos de este nivel empiezan a abordar aspectos del pensamiento crítico, como la evaluación de la razonabilidad de un resultado mediante juicios cualitativos cuando no es posible realizar cálculos a partir de la información dada. Pueden seleccionar e integrar distintas representaciones de la información, incluidas las simbólicas o gráficas, relacionándolas directamente con aspectos de situaciones del mundo real. En este nivel, los alumnos también pueden construir y comunicar explicaciones y argumentos basados en sus interpretaciones, razonamientos y metodología.

**Nivel 3 (482)**

En el Nivel 3, los alumnos son capaces de elaborar estrategias de solución, incluidas aquellas que requieren toma de decisiones secuenciales o flexibilidad en la comprensión de conceptos familiares. En este nivel, los alumnos comienzan a utilizar habilidades de pensamiento computacional para desarrollar su estrategia de solución. Son capaces de resolver tareas que requieren la realización de varios cálculos diferentes pero rutinarios que no están todos claramente definidos en el enunciado del problema. Pueden utilizar la visualización espacial como parte de una estrategia de solución o determinar cómo utilizar una simulación para recopilar datos apropiados para la tarea. Los alumnos de este nivel pueden interpretar y utilizar representaciones basadas en distintas fuentes de información y razonar directamente a partir de ellas, incluida la toma de decisiones condicional utilizando una tabla de dos direcciones. Suelen mostrar cierta habilidad para manejar porcentajes, fracciones y números decimales, y para trabajar con relaciones proporcionales.

**Nivel 2 (420)**

En el Nivel 2, los alumnos pueden reconocer situaciones en las que necesitan diseñar estrategias sencillas para resolver problemas, incluida la realización de simulaciones sencillas con una variable como parte de su estrategia de solución.

**Nivel de competencia y puntuación de corte inferior**

**Descripción de la tarea**

Pueden extraer información relevante de una o más fuentes que utilizan modos de representación ligeramente más complejos, como tablas de dos vías, gráficos o representaciones bidimensionales de objetos tridimensionales. Los alumnos de este nivel demuestran una comprensión básica de las relaciones funcionales y pueden resolver problemas que impliquen relaciones sencillas. Son capaces de hacer interpretaciones literales de los resultados.

**Nivel 1a (358)**

En el Nivel 1a, los alumnos pueden responder a preguntas que implican contextos sencillos en los que toda la información necesaria está presente y las preguntas están claramente definidas. La información puede presentarse en diversos formatos sencillos y los alumnos pueden tener que trabajar con dos fuentes simultáneamente para extraer la información pertinente. Son capaces de llevar a cabo procedimientos sencillos y rutinarios siguiendo instrucciones directas en situaciones explícitas, que en ocasiones pueden requerir múltiples iteraciones de un procedimiento rutinario para resolver un problema. Pueden realizar acciones que son obvias o que requieren una síntesis mínima de información, pero en todos los casos las acciones se derivan claramente de los estímulos dados. Los alumnos de este nivel pueden emplear algoritmos básicos, fórmulas, procedimientos o convenciones para resolver problemas que, en la mayoría de los casos, implican números enteros.

**Nivel 1b (295)**

En el Nivel 1b, los alumnos son capaces de responder a preguntas que implican contextos de fácil comprensión en los que toda la información necesaria se ofrece claramente en una representación sencilla (es decir, tabular o gráfica) y, en caso necesario, reconocer cuándo alguna información es ajena y puede ignorarse con respecto a la pregunta concreta que se plantea. Es capaz de realizar cálculos sencillos con números enteros a partir de instrucciones claramente prescritas y definidas en un texto breve y sintácticamente sencillo.

**Nivel 1c (233)**

En el Nivel 1c, los alumnos son capaces de responder a preguntas en contextos

Nivel de competencia  
y puntuación de corte  
inferior

Descripción de la tarea

de fácil comprensión en los que toda la información relevante se ofrece de forma clara, en un formato sencillo y familiar (por ejemplo, una tabla pequeña o un dibujo) y se define en un texto muy breve y sintácticamente sencillo. Es capaz de seguir unas instrucciones claras que describen un único paso u operación.

NOTA: Para alcanzar un determinado nivel de competencia, el alumno debe responder correctamente a la mayoría de los ítems de ese nivel. Los alumnos se clasificaron en niveles de competencia matemática en función de sus puntuaciones. Las puntuaciones se expresan en una escala de 0 a 1.000.

Fuente: Organización para la Cooperación y el Desarrollo Económico (OCDE), Programa para la Evaluación Internacional de Alumnos (PISA), 2022.

## **Recomendaciones generales**

El análisis realizado en este informe sobre las pruebas estandarizadas sumativas aplicadas por el MEP en sexto grado y en el 2023, evidencia la existencia de una pobreza de estas evaluaciones que requiere atención y revisión por parte del MEP, independientemente de si son instrumentos con fines diagnósticos o de certificación. Sufrimos un problema serio de validez en los instrumentos existentes y se presenta un vacío en torno a otros que requerimos y que no existen. Esta situación sucede precisamente en el momento que más necesitamos contar con datos válidos y confiables para la toma de decisiones.

Con el fin de aportar a dicha revisión se presentan a continuación un conjunto de recomendaciones:

Ante los crecientes procesos de desigualdad en oportunidades educativas que no han podido ser solventados por el Estado, se debe utilizar la macro evaluación con propósitos diagnósticos y formativos si la intención es corregir y mejorar, en términos de excelencia y equidad, pues como ya se ha mostrado a lo largo de este trabajo, la macro evaluación con propósitos de promoción para los estudiantes no es suficiente para lograr estos propósitos de manera óptima.

Partiendo de lo establecido en los programas de estudio, se recomienda utilizar un conjunto de estándares educativos con sus respectivos niveles de desempeño, debidamente discutidos y consensuados con los diversos actores educativos, para contar con un punto de partida más claro y consistente en torno a las tablas de especificaciones que deberían ser la base para las pruebas estandarizadas de diagnóstico y formativas, y también para las pruebas de certificación. Al respecto es necesario indicar que ya se avanzó en un primer esfuerzo para la creación de esos estándares en español, matemáticas y ciencias, teniendo como insumos a los actuales programas de estudio (Montero-Rojas, 2023).

La validez de contenido (alineación curricular) en términos de una cobertura adecuada de los aprendizajes al finalizar la primaria y la secundaria, debe ser una de las principales características que debe buscarse en las metodologías para la construcción y validación de dichas pruebas.

Aunado a lo anterior, se debe promover una sana cultura evaluativa donde se enfatice el papel formativo de la evaluación educativa estandarizada, para diagnosticar, aprender, corregir y mejorar, desproveyéndole el carácter punitivo que han tenido las pruebas estandarizadas para los estudiantes costarricenses durante los últimos 35 años, considerando igualmente el cambio contextual que la sociedad costarricense ha experimentado en dicho período, donde se han acrecentado velozmente las desigualdades en oportunidades educativas.

En este contexto, se debe reforzar la idea de la evaluación como bien público, al igual que debe ser la educación.

Por otra parte, se sugiere tomar como referencia inicial los niveles de desempeño de las pruebas ERCE y asegurarse de que estén presentes en los ítems de las pruebas estandarizadas de español y matemática de sexto grado, pues ellos representan los estándares mínimos que deberían evaluarse, dado que han sido consensuados con todos los países de América Latina, a partir de sus marcos curriculares.

En el caso de la secundaria, las pruebas de PISA se pueden utilizar como referencia para la construcción de pruebas de diagnóstico a partir de noveno año, si bien la incorporación de los niveles de desempeño más altos de esta prueba internacional debe ser gradual y apoyada con propuestas pedagógicas adecuadas.

En el caso de las evaluaciones estandarizadas con propósitos formativos o de diagnóstico, se hace necesario incorporar, a nivel metodológico, el uso de diseños de *matrix sampling*, tal se hace en las pruebas internacionales PISA y ERCE, y también como sucede en muchas de las pruebas nacionales alrededor del mundo, para estar en capacidad de cubrir todos los aprendizajes relevantes, y generar niveles de desempeño válidos, sin que ello implique muchas horas de aplicación para los estudiantes. Por cierto, Uruguay y Chile, los dos países líderes en América Latina en las pruebas PISA, implementan estos diseños en sus pruebas estandarizadas nacionales.

Conociendo que el uso correcto del lenguaje y la comunicación está en la base de los aprendizajes para todas las demás asignaturas, es esencial incorporar evaluaciones estandarizadas con un nivel adecuado de calidad para medir los otros tres pilares de la comunicación: escucha, escritura y habla, además de la comprensión lectora.

Refiriéndose específicamente a la evaluación de la comprensión lectora, existen a nivel internacional diversas fórmulas que se han propuesto para valorar la “readability” (legibilidad o “lecturabilidad”) de un texto con una métrica cuantitativa. La mayoría de ellas también ofrecen

equivalencias en torno a los niveles de lectura según el grado escolar (Jiménez Pérez, López Martínez & Corbalán Berná, 2024; Lennon & Burdick, 2014).

Se recomienda entonces utilizar estas medidas para adecuar los textos de las pruebas comprensión lectora al nivel de las competencias lectoras que se espera de los estudiantes al final de la primaria y la secundaria.

Asimismo, se recomienda utilizar el listado de características óptimas de una prueba estandarizada educativa, mostrado en el cuadro 2 en este documento, cuando el propósito del instrumento estandarizado es la mejora educativa, tanto en términos de excelencia como de equidad.

Además, en un contexto macroestructural y considerando que la pobreza de los aprendizajes y la pobreza de las evaluaciones son problemas interconectados que impiden el desarrollo educativo en Costa Rica, se considera que, para lograr una mejora significativa, se requiere un enfoque integral que aborde las causas subyacentes, mejore la calidad de las evaluaciones y reduzca la desigualdad. Por tanto, se recomienda también:

- Crear un centro independiente de macro evaluación educativa para evaluar de manera válida todos los componentes del sistema, teniendo como propósito el mejoramiento educativo. Uno de sus propósitos sería cambiar la cultura evaluativa, promoviendo a la evaluación como herramienta para la mejora continua.
- Atender las inequidades: Implementar políticas para reducir la desigualdad social y mejorar las oportunidades educativas para todas las personas.
- Para lo anterior, y tal como se indica en la publicación *La encrucijada de la educación en América Latina y el Caribe: Informe regional de monitoreo ODS4-Educación 2030* (UNESCO, CEPAL & UNICEF, 2022) ya existen pistas para orientar el rumbo. Se destaca la importancia crucial de la mejora de las condiciones de vida de la población como primer factor que influye en los aprendizajes: “...las políticas educativas son centrales y pueden hacer grandes diferencias, pero no bastan si no cuentan con una mirada más amplia del desarrollo económico con inclusión social” (p.211).

## Referencias bibliográficas

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Araki, S., & Kariya, T. (2022). Credential inflation and decredentialization: Re-examining the mechanism of the devaluation of degrees. *European Sociological Review*, 38(6), 904–919. <https://doi.org/10.1093/esr/jcac004>
- Azevedo, João Pedro, Diana Goldemberg, Silvia Montoya, Reema Nayar, Halsey Rogers, Jaime Saavedra, and Brian William Stacy (2021). “Will Every Child Be Able to Read by 2030? Defining Learning Poverty and Mapping the Dimensions of the Challenge.” Working Paper. Washington, DC: World Bank. doi:10.1596/1813-9450-9588.
- Banco Mundial (2019). Pobreza en el aprendizaje: una tarea pendiente en Latinoamérica y el Caribe. <https://www.bancomundial.org/es/news/feature/2019/10/21/pobrezaaprendizaje-latinoamerica-caribe>.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
- Esquivel Alfaro, J. M., Montero Rojas, I., Sosa Jara, D., Hernández Rodríguez, A. C., Corella Espinoza, M. I., & Fallas Monge, J. (2006). Evaluación externa de las pruebas nacionales de bachillerato de la educación media. Universidad de Costa Rica, Facultad de Educación, Sistema de Estudios de Posgrado, Programa de Posgrado en Evaluación Educativa.
- Jiménez Pérez, E. del P., López Martínez, O., & Corbalán Berná, J. (2024). Algoritmo LECOLE: legibilidad, lecturabilidad y comprensibilidad en español. *XLingua*, 17(1), 12. <https://doi.org/10.18355/XL.2024.17.01.12>
- Lennon, C., & Burdick, H. (2014). *The Lexile® framework as an approach for reading measurement and success*. MetaMetrics, Inc.
- Linacre, J. M. (2002). *Winsteps Rasch measurement computer program*. Winsteps.com.
- Ministerio de Educación Pública. Dirección de Gestión y Evaluación de la Calidad. (2024a, junio). *Prueba Nacional Estandarizada. Aplicación sumativa. Primaria 2023: Informe para el docente* [Informe]. <https://dgec.mep.go.cr/wp-content/uploads/2024/06/Informe-docente-Sumativa-PRIMARIA.pdf>
- Ministerio de Educación Pública. Dirección de Gestión y Evaluación de la Calidad. (2024b, junio). *Prueba Nacional Estandarizada. Aplicación sumativa. Secundaria académico diurno 2023: Informe para el docente* [Informe]. <https://dgec.mep.go.cr/wp->

[content/uploads/2024/06/Informe-docente-Sumativa-SECUNDARIA-ACADEMICA-DIURNO.pdf](https://content/uploads/2024/06/Informe-docente-Sumativa-SECUNDARIA-ACADEMICA-DIURNO.pdf)

- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.  
<https://doi.org/10.3102/0013189X018002005>
- Montero-Rojas, E. (2023). *Importancia de los estándares educativos para evaluar el progreso de la educación costarricense*. Ponencia para el Noveno Informe del Estado de la Educación. Programa Estado de la Nación. Consejo Nacional de Rectores (CONARE).
- Montero-Rojas, E. (2013). *Referentes conceptuales y metodológicos sobre la noción moderna de validez de instrumentos de medición: implicaciones para el caso de personas con necesidades educativas especiales*. *Actualidades en Psicología*, Vol. 27, núm.114. San José, Costa Rica: Universidad de Costa Rica.  
DOI: <http://dx.doi.org/10.15517/ap.v27i114.7900>
- Montero-Rojas, E., & Rojas-León, S. (2012). *El modelo de Rasch: Una herramienta esencial para la evaluación educativa de gran escala*. Aporte Especial. Cuarto Informe Del Estado De La Educación.
- Moreira-Mora, T. E., Alfaro-Rojas, L., Brizuela-Rodríguez, A., Chacón-Vega, C., Gómez-González, E., Jiménez-Alfaro, K., Jiménez-Segura, F., Mena-Castillo, P., Montero-Rojas, E., Picado-Barrantes, H., Rojas-Rojas, G., Rojas-Torres, L., Smith-Castro, V., Solórzano-Salas, M. J., & Villarreal-Galera, M. P. (2022). *Cuaderno 9: Estándares de calidad para pruebas estandarizadas de alto impacto en el contexto académico y profesional costarricense*. Instituto de Investigaciones Psicológicas. Universidad de Costa Rica.  
<https://www.iip.ucr.ac.cr/es/publicaciones/series-cuadernos-metodologicos/cuaderno-9-estandares-de-calidad-para-pruebas>
- Muñiz, J. (2003). *Teoría clásica de los tests* (2ª ed.). Madrid: Pirámide.
- Murillo Rojas, M. & Barquero Mejías, K. (2025). *Análisis de la Prueba Estandarizada del componente Español, año 2023, Segundo Ciclo de EGB: Aportes para la discusión nacional*. Ponencia para el Décimo Informe del Estado de la Educación. Programa Estado de la Nación. Consejo Nacional de Rectores (CONARE).
- Murillo Rojas, M., Araya Ramírez, J., & Barquero Mejías, K. (2023). *Situación de la pobreza de los aprendizajes en la población estudiantil de 10 años en el período 2019-2022*. Programa Estado de la Nación (PEN), Noveno Informe Estado de la Educación, Capítulo 2. San José, Costa Rica: PEN. <https://hdl.handle.net/20.500.12337/8519>
- OECD (2023), *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/53f23881-en>.
- Picado Arce, K., Conejo Bolaños, L. D., & Montero Rojas, E. (2025). Más allá de la calidad educativa: excelencia y equidad como conceptos clave para evaluar el sistema

educativo costarricense. *Actualidades Investigativas En Educación*, 25(2), 1–30.  
<https://doi.org/10.15517/aie.v25i2.62369>

Programa Estado de la Nación (2023). *Noveno Informe Estado de la Educación*. San José: Programa Estado de la Nación.

UNESCO (2021). *Estudio Regional Comparativo y Explicativo (ERCE 2019): Reporte nacional de resultados, Costa Rica*. UNESCO Office Santiago and Regional Bureau for Education in Latin America and the Caribbean, Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. <https://unesdoc.unesco.org/ark:/48223/pf0000380244>

UNESCO Institute for Statistics. (2020). *Learning Poverty: Measurement and Simulations (WG/GAML/7)*. UNESCO. [https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2020/10/wg-gaml-7-learning-poverty\\_measurement-and-simulations.pdf](https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2020/10/wg-gaml-7-learning-poverty_measurement-and-simulations.pdf)

UNESCO, CEPAL, & UNICEF. (2022). *La encrucijada de la educación en América Latina y el Caribe: Informe regional de monitoreo ODS4-Educación 2030*. UNICEF.

UNESCO. (2020). *Análisis curricular: Estudio Regional Comparativo y Explicativo (ERCE 2019) Costa Rica: Documento nacional de resultados*.

World Bank, UNESCO, UNICEF, USAID, FCDO, & Bill & Melinda Gates Foundation. (2022). *The state of global learning poverty: 2022 update (Conference edition)*. <https://www.unicef.org/media/122921/file/StateofLearningPoverty2022.pdf>

Zumbado M., Hernández L., Poveda R., Ramírez L.F. & Sánchez A. (2025). *Desarrollo de Habilidades Fundamentales en Matemáticas: Resultados PISA, programas de estudio vigentes y evaluaciones nacionales*. Ponencia para el Décimo Informe del Estado de la Educación. Programa Estado de la Nación. Consejo Nacional de Rectores (CONARE).