

COMPETENCIA
LECTURA
MATEMÁTICA
CIENCIA
SOCIOEMOCIONAL
MEDICIÓN
DESEMPEÑO
PSICOMETRÍA
APRENDIZAJE
EDUCACIÓN
CALIDAD

Reporte técnico de la Evaluación Nacional de Logros de Aprendizaje de Estudiantes (ENLA) 2023



PERÚ

Ministerio
de Educación



BICENTENARIO
PERÚ
2024

COMPETENCIA
LECTURA
MATEMÁTICA
CIENCIA
SOCIOEMOCIONAL
MEDICIÓN
DESEMPEÑO
PSICOMETRÍA
APRENDIZAJE
EDUCACIÓN
CALIDAD

Reporte técnico de la Evaluación Nacional de Logros de Aprendizaje de Estudiantes (ENLA) 2023



PERÚ

Ministerio
de Educación



BICENTENARIO
PERÚ
2024



PERÚ

Ministerio
de Educación

Morgan Niccolo Quero Gaimé

Ministro de Educación del Perú

Cecilia del Pilar García Díaz

Viceministra de Gestión Institucional

María Esther Cuadros Espinoza

Viceministra de Gestión Pedagógica

Nestor Alfonso Supanta Velásquez

Secretario de Planificación Estratégica

Diego Arturo Luna Vera Tudela

Jefe de la Oficina de Medición de la Calidad de los Aprendizajes

**Reporte técnico de la Evaluación Nacional de Logros de Aprendizaje
(ENLA) 2023**

Responsables del documento

Andrés Alberto Burga León (Coordinador)

Tania Pacheco Valenzuela

Yoni Arámbulo Mogollón

Rosina Vanessa Sánchez Jiménez

Luis Mejía Campos

Gabriela Santibáñez Rojas

Marco Aurelio Carranza Rodríguez

Christian Alexis Moreau Almaraz

Wilder Iván Lázaro Illatopa

Sandra Elizabeth Flores Ari

Victor Wilfredo Salazar Oré

Diego Orlando Hermoza Cardenas

Esta publicación es el producto final del esfuerzo institucional de la UMC por medio de sus diferentes equipos de especialistas.

©Ministerio de Educación del Perú, 2024

Calle Del Comercio 193, San Borja

Lima, Perú

Teléfono: (511) 615-5800

www.minedu.gob.pe

Todos los derechos son reservados. Prohibida la reproducción de este libro por cualquier medio, total o parcialmente, sin permiso expreso.

Índice

Introducción	7
Capítulo 1: Diseño y construcción de las pruebas aplicadas	9
1.1 Construcción de los ítems y las pruebas	12
1.2 Procesos de validación de los instrumentos de evaluación	15
1.2.1 Validación de expertos de la UMC	16
1.2.2 Validación de campo: aplicación piloto.....	17
1.2.3 Revisión de ítems y consulta con la población objetivo.....	17
1.2.4 Análisis de las propiedades psicométricas y de la distribución de los ítems en la escala.....	18
1.2.5 Validación de expertos de otras oficinas del Minedu.....	18
1.2.6 Adaptación de instrumentos para estudiantes con discapacidad	19
1.2.7 Aplicación definitiva.....	22
1.3 Ensamblaje de las pruebas mediante el diseño bloques	23
1.3.1 Diseño de bloques de la ENLA 2023	24
Capítulo 2: Población a evaluar	44
2.1 Población objetivo	45
2.2 Marco censal y marco muestral	45
2.3 Diseño muestral.....	46
2.4 Cobertura en la ENLA 2023.....	49
2.5 Pesos muestrales ENLA 2023	53
Capítulo 3: Operativo de campo	56
3.1 Organización territorial del operativo	58
3.2 Conformación de la Red Administrativa de aplicación	59
3.3 Proceso de selección y capacitación del personal de la RA.....	63
3.4 Características de los instrumentos y documentos utilizados	69
3.5 Proceso de aplicación	71
Capítulo 4: Procesamiento de datos y análisis psicométrico	77
4.1 Gestión y depuración de datos	78
4.1.1 Gestión física	78
4.1.2 Digitalización y captura de datos	79
4.1.3 Depuración de datos.....	80

4.2	Estrategias de análisis psicométrico.....	81
4.2.1	Modelo Rasch para ítems dicotómicos.....	81
4.2.2	Modelo Rasch de crédito parcial.....	84
4.2.3	Análisis de ítems.....	87
4.2.4	Evidencias de validez vinculadas a la estructura interna.....	89
4.2.5	Confiabilidad y consistencia de la clasificación.....	90
4.3	Resultados del análisis psicométrico en Lectura, Matemática, y Ciencias Sociales.....	91
4.3.1	Calibración de los ítems.....	91
4.3.2	Evidencias de validez vinculadas a la estructura interna.....	113
4.3.3	Confiabilidad y consistencia de la clasificación.....	114
4.4	Equiparación de medidas de Educación Básica Regular (EBR).....	115
4.5	Estrategias de análisis psicométrico en educación intercultural bilingüe (EIB).....	126
4.5.1	Modelo TRI logístico de tres parámetros.....	126
4.5.2	Análisis de ítems.....	128
4.5.3	Evidencias de validez vinculadas a la estructura interna de las pruebas.....	129
4.5.4	Confiabilidad y consistencia de la clasificación.....	129
4.6	Resultados del análisis psicométrico en Lectura castellano como segunda lengua EIB.....	132
4.6.1	Calibración de los ítems.....	132
4.6.2	Evidencias de validez vinculadas a la estructura interna.....	136
4.6.3	Confiabilidad y consistencia de la clasificación.....	137
4.7	Niveles de logro y preparación de resultados.....	138
4.8	Análisis psicométricos para habilidades socioemocionales.....	141
4.8.1	Calibración de los ítems.....	142
4.8.2	Estructura de umbrales.....	143
4.8.3	Evidencias de validez vinculadas a la estructura interna.....	144
4.8.4	Confiabilidad.....	144
Capítulo 5: Procesamiento de datos de Factores Asociados.....		145
5.1	Elaboración de las preguntas del cuestionario de factores asociados.....	146
5.1.1	Algunos conceptos claves.....	147
5.1.2	Procedimiento para la elaboración de cuestionarios.....	148

5.2	Elaboración de análisis descriptivo e imputación de datos.....	151
5.3	Validez y confiabilidad de los índices y escalas de factores asociados	151
5.3.1	Evidencias de validez basadas en la estructura interna	152
5.3.2	Evidencias de Confiabilidad	161
5.3.3	Elaboración de puntajes	170
	Referencias	171
	Anexos	178

Introducción

La Oficina de Medición de la Calidad de los Aprendizajes (UMC), del Ministerio de Educación, según sus funciones, diseña e implementa las evaluaciones de logros de aprendizaje de los estudiantes de Educación Básica a nivel nacional. Estas evaluaciones proporcionan información valiosa sobre los logros y desafíos en los aprendizajes, con el propósito de contribuir con evidencias en el diseño de planes y estrategias orientadas a la mejora de los aprendizajes en las áreas y competencias evaluadas. Este aporte se da tanto a nivel del sistema educativo como en el ámbito de los actores: padres de familia, directivos y docentes de las instituciones educativas, especialistas y directores de las Unidades de Gestión Educativa Local (UGEL), Direcciones y Gerencias Regionales de Educación (DRE/GRE), especialistas y funcionarios del Minedu.

En este contexto, el Minedu publicó la Resolución Ministerial N.º 190-2023-MINEDU que aprueba la Evaluación Nacional de Logros de Aprendizaje – ENLA 2023, con el objetivo de obtener un diagnóstico actualizado de los aprendizajes de los estudiantes. Dicha evaluación se aplicó de manera presencial a una muestra representativa de instituciones educativas a nivel nacional.

La ENLA 2023 evaluó a estudiantes de 2.º grado de primaria en Comunicación (Lee diversos tipos de textos escritos en su lengua materna) y Matemática (Resuelve problemas de cantidad); en 4.º grado de primaria en Comunicación (Lee diversos tipos de textos escritos en su lengua materna) y en Matemática (todas las competencias); y en 2.º grado de secundaria en Comunicación (Lee diversos tipos de textos escritos en su lengua materna), Matemática (todas las competencias) y Ciencias Sociales (Construye interpretaciones históricas, Gestiona responsablemente el espacio y el ambiente, y Gestiona responsablemente los recursos económicos). Asimismo, la ENLA 2023 incorporó una prueba de carácter censal a estudiantes de 4.º grado de primaria en Comunicación (Lectura en castellano como segunda lengua) en las instituciones educativas de Educación Intercultural Bilingüe (EIB) de Fortalecimiento Cultural y Lingüístico.

La ENLA 2023 incluyó la aplicación de cuestionarios dirigidos a directores, docentes, familias y estudiantes para entender mejor los factores que influyen en el éxito educativo. Estos cuestionarios ayudan a identificar las variables que pueden favorecer o dificultar el logro de los aprendizajes. Además, para evaluar las habilidades socioemocionales, se realizó un cuestionario específico aplicado a estudiantes de 2.º grado de secundaria y 6.º grado de primaria, logrando proporcionar una visión integral de su bienestar emocional y social.

En el primer capítulo de este documento, se describe el proceso de construcción de los instrumentos aplicados. Se señalan los criterios que se siguen en la elaboración de los ítems, el recojo de evidencias de validez vinculadas al contenido de los ítems y el proceso de las aplicaciones piloto y definitiva.

En el segundo capítulo, se describen la población objetivo y el marco muestral utilizado. Además, se señala cómo fueron definidos los tamaños y pesos muestrales utilizados en la estimación de los diversos parámetros. Este capítulo incluye el análisis de la cobertura.

El tercer capítulo aborda los aspectos relacionados con el operativo de campo. Describe cómo se organizó la red administrativa, cómo fueron seleccionados y capacitados los aplicadores, cuáles fueron los instrumentos utilizados y cómo se recolectaron los datos en el contexto de las evaluaciones del 2023.

El cuarto capítulo describe el proceso de gestión y depuración de datos, así como el modelo Rasch utilizado para los análisis psicométricos. Se pone énfasis en cómo se analizaron los ítems y cómo se obtuvieron evidencias de confiabilidad y validez. Asimismo, se describe cómo se realizó el proceso de equiparación de medidas y la forma en que se establecieron los niveles de logros a partir de los puntos de corte.

Finalmente, el quinto capítulo describe el proceso relacionado con la elaboración de los cuestionarios y el procesamiento de datos de factores asociados. Se considera la elaboración de análisis descriptivo e imputación de datos y los aspectos vinculados a la validez y confiabilidad.

Diseño y construcción de las pruebas aplicadas

Capítulo 1

La Oficina de Medición de la Calidad de los Aprendizajes (UMC) cuenta con un Equipo de Evaluación conformado por distintos subequipos organizados según especialidades: Comunicación, Matemática, Lenguas originarias, Ciudadanía y Ciencias Sociales, Ciencia y Tecnología. Además, cuenta con un equipo de Habilidades Socioemocionales. Los especialistas de estos equipos son los encargados de diseñar y construir las pruebas y demás instrumentos, para las diversas evaluaciones y estudios que están bajo la responsabilidad de la UMC como instancia técnica del Ministerio de Educación del Perú (Minedu). Asimismo, la oficina cuenta con un subequipo de Atención a la Diversidad que orienta y contribuye a la aplicación del diseño universal de evaluación para la construcción de todas las pruebas, así como en la adaptación de estas y de los procedimientos de aplicación a las características particulares de los estudiantes con discapacidad incluidos en las escuelas de Educación Básica Regular (EBR), con la finalidad de que puedan participar en las evaluaciones de forma autónoma.

Cada año, las pruebas que se aplican en el marco de las evaluaciones nacionales de logros de aprendizaje con alcance censal o muestral son especificadas mediante una resolución ministerial publicada durante los primeros meses del año en el diario oficial El Peruano. Así, en el 2023, mediante R.M. 190-2023-Minedu, se dispuso la implementación de la Evaluación Nacional de Logros de Aprendizaje de Estudiantes (ENLA), tal como se detalla en la tabla 1.1.

Tabla 1.1 Evaluaciones muestrales y censal implementadas por la UMC en el 2023

Nivel	Grado	Área	Competencia(s) evaluada(s)	Denominación de la prueba
Primaria	2.º	Comunicación	Lee diversos tipos de textos escritos en su lengua materna.	Lectura
		Matemática	Resuelve problemas de cantidad.	Matemática

Nivel	Grado	Área	Competencia(s) evaluada(s)	Denominación de la prueba
Primaria	4.º	Comunicación	Lee diversos tipos de textos escritos en su lengua materna.	Lectura
		Matemática	Resuelve problemas de cantidad.	Matemática
			Resuelve problemas de regularidad, equivalencia y cambio.	
			Resuelve problemas de gestión de datos e incertidumbre.	
	Resuelve problemas de forma, movimiento y localización.			
4.º EIB	Castellano como segunda lengua	Lee diversos tipos de textos escritos en castellano como segunda lengua.	Lectura	
Secundaria	2.º	Comunicación	Lee diversos tipos de textos escritos en su lengua materna.	Lectura
		Matemática	Resuelve problemas de cantidad.	Matemática
			Resuelve problemas de regularidad, equivalencia y cambio.	
			Resuelve problemas de gestión de datos e incertidumbre.	
	Resuelve problemas de forma, movimiento y localización.			
	Ciencias Sociales	Construye interpretaciones históricas.	Ciencias Sociales	
		Gestiona responsablemente el espacio y el ambiente.		
		Gestiona responsablemente los recursos económicos.		
Habilidades Socioemocionales	Toma de decisiones. Resolución de problemas.	Habilidades Socioemocionales		

1.1 Construcción de los ítems y las pruebas

Los procesos del diseño, así como la construcción de los ítems y las pruebas, tienen como referente fundamental el Currículo Nacional de Educación Básica (CNEB, aprobado mediante Resolución Ministerial N.º 281-2016-MINEDU), y siguen los principios de validez, confiabilidad e imparcialidad. Por un lado, tomar como referente el currículo involucra la evaluación de las competencias establecidas en este documento desde la perspectiva del enfoque propio de cada área: el enfoque comunicativo (y sociocultural) del área de Comunicación, el enfoque de resolución de problemas del área de Matemática y el enfoque de ciudadanía activa del área de Ciencias Sociales. Asimismo, seguir los principios de validez, confiabilidad e imparcialidad implica que los instrumentos de evaluación deben recoger información de los estudiantes de tal manera que se pueda estimar de forma fiable su nivel de logro en las competencias evaluadas, y que dicha información pueda ser usada para los fines propios del proceso educativo (American Educational Research Association, American Psychological Association y National Council of Measurement in Education, 2014). Además, los instrumentos deben reflejar una concepción inclusiva de la educación, conforme a los lineamientos de la política educativa nacional (Ministerio de Educación, 2015).

En las evaluaciones nacionales de logros de aprendizaje, se emplean dos formatos de ítems: de opción múltiple y de respuesta construida. A continuación, se describe cada uno de ellos.

Ítem de opción múltiple. El formato de opción múltiple es el más común. Presenta en su estructura un estímulo, enunciado o pregunta y tres o cuatro alternativas de respuesta, entre las cuales el estudiante debe identificar la única que es correcta. Esto implica que dos o tres alternativas funcionan como distractores (respuestas incorrectas). En ciertos ítems, alguno de los distractores puede constituir una respuesta intermedia o de logro parcial y, por lo tanto, se modela como un ítem de crédito o puntaje parcial. En las evaluaciones implementadas por la UMC, las pruebas de 2.º grado de primaria en su mayoría están conformadas por ítems que presentan tres alternativas para evitar que el estudiante, debido a su edad o al nivel de desarrollo de su competencia lectora, haga un mayor esfuerzo cognitivo en la lectura del ítem que en su solución. Para los demás grados, las pruebas están conformadas por ítems que presentan cuatro alternativas.

La construcción del enunciado y de las alternativas de un ítem de opción múltiple sigue las pautas ofrecidas por Moreno, Martínez y Muñiz (2004). De acuerdo con lo propuesto por estos autores, se considera lo siguiente:

Respecto de la construcción del ítem en general

- Debe haber correspondencia entre el ítem y el indicador que evalúa.

- La complejidad de la tarea solicitada debe ser adecuada para el estudiante según el grado en que se encuentra.
- El ítem debe estar libre de todo sesgo sociocultural, socioeconómico, de género y de religión.
- El ítem no debe estar relacionado con ningún otro ítem de la prueba; es decir, que la respuesta correcta no puede inferirse a partir de otro ítem.
- El ítem debe ser fácil de entender; es decir, la persona evaluada debe comprender con claridad qué tarea debe realizar (al margen de que pueda realizarla).
- El ítem debe tener elementos visuales y gráficos solo cuando sean necesarios para responderlo.
- El ítem debe tener una alta legibilidad material; es decir, que el tamaño y el tipo de fuente, los espacios entre caracteres y entre líneas, el contraste, entre otras características, deben ser los adecuados.

Respecto de la construcción del enunciado

- Lo central debe expresarse en el enunciado. Cada opción es un complemento de este.
- El enunciado debe concordar gramaticalmente con todas las alternativas.
- La sintaxis o estructura gramatical debe ser clara y precisa; es decir, que el enunciado debe estar claramente redactado. Se deben evitar enunciados demasiado escuetos, profusos, ambiguos o confusos.
- En lo posible, debe evitarse redactar el enunciado de forma negativa. Si hubiera necesidad de usar una expresión negativa, esta debe resaltarse.
- El enunciado debe estar libre de elementos irrelevantes e innecesarios.

Respecto de la construcción de las alternativas

- La alternativa correcta debe ser solo una y estar acompañada por distractores plausibles.
- A lo largo de la prueba, la alternativa correcta debe estar distribuida entre las distintas ubicaciones.
- Las alternativas deben ser, preferiblemente, tres o cuatro.
- Las alternativas deben disponerse, de preferencia, de forma vertical.

- Las alternativas de cada ítem deben organizarse de acuerdo con un criterio; por ejemplo, del número menor al número mayor.
- Las alternativas deben ser autónomas entre sí, sin solaparse ni referirse unas a otras.
- Se debe evitar las opciones “Todas las anteriores” y “Ninguna de las anteriores”.
- Ninguna alternativa debe destacar entre las demás, ni en contenido ni en apariencia.
- Las alternativas deben estar libres de palabras, frases o imágenes irrelevantes.
- Se deben emplear como distractores los errores posibles y más frecuentes en la resolución del ítem.

Ítem de respuesta construida. Este formato de ítem se compone de un estímulo y de un enunciado que demandan al estudiante proporcionar, desarrollar o crear una respuesta particular. El estímulo consiste en una situación contextualizada o un texto (en el caso de la prueba de Lectura) que le da un marco tanto al ítem como a la respuesta esperada. Este tipo de ítem también puede incluir instrucciones o especificaciones para ser respondido.

La inclusión de un ítem de respuesta construida responde a la necesidad de explorar tanto los procedimientos como las habilidades complejas de argumentación, evaluación, juicio crítico y toma de decisiones razonada involucradas en la resolución de ciertas tareas. Si bien se pueden formular ítems cerrados de opción múltiple para evaluar estas habilidades complejas, los ítems de respuesta construida permiten medirlas en toda su dimensión y con mayor profundidad.

Las respuestas de los estudiantes a un ítem de respuesta construida son codificadas por expertos en el área, especialmente capacitados para este fin. Los codificadores pasan por un riguroso proceso de capacitación que les permite apropiarse de la lógica de los ítems, los criterios de codificación y los posibles tipos de respuestas. Tanto para el proceso de capacitación como para la codificación en sí, se utiliza un manual de codificación que especifica el indicador y el descriptor que el ítem evalúa, estableciendo los tipos de respuesta (de crédito total, crédito parcial o sin crédito), así como los criterios de codificación para cada uno ellos. Además, este manual brinda al menos cuatro ejemplos de respuestas reales de los estudiantes para cada uno de los criterios establecidos.

Las respuestas de crédito total son aquellas que muestran el logro completo de los procedimientos y de las habilidades implicadas en la resolución de la tarea. En este proceso, un estudiante puede utilizar diferentes procedimientos. Por su parte, las respuestas de crédito parcial son aquellas que reflejan un logro parcial del estudiante

en la resolución de la tarea, mientras que las respuestas sin crédito evidencian una comprensión errada y, por lo tanto, no responden a los criterios de codificación establecidos.

En las pruebas de las evaluaciones nacionales de logros de aprendizaje, hay una menor cantidad de ítems de respuesta construida en relación con los ítems de opción múltiple. Esto se debe a que el procesamiento de las respuestas de este tipo de ítems es complejo y, en aplicaciones masivas como la ENLA 2023, demanda muchos recursos y tiempo. Finalmente, se utiliza el formato tipo Likert en los instrumentos de habilidades socioemocionales.

Ítem tipo Likert. Este formato de ítem se compone usualmente de un enunciado y de diferentes alternativas de respuesta que conforman una escala ordinal (p. ej.: nunca, casi nunca, a veces, casi siempre, siempre). Este tipo de ítem surge con la publicación de Rensis Likert en 1932, en la cual se presenta un tipo de instrumento para la medición de las actitudes (Likert, 1932). Estos ítems son preguntas individuales que pueden ser colocadas en un instrumento sin intención de combinarlas en un puntaje compuesto (composite score), y ser analizadas a nivel ordinal. También permiten diseñar escalas, en las que se combinan las respuestas de los distintos ítems para representar la variable o el constructo a medir. El tipo de escala usado en un ítem depende del constructo que se está tratando de medir, siendo las anclas aquellas “palabras” utilizadas para identificar cada categoría de respuesta. Es común el uso de escalas de acuerdo, en que las palabras ancla suelen ser Totalmente de acuerdo, De acuerdo, Ni de acuerdo ni en desacuerdo, En desacuerdo, Totalmente en desacuerdo. En la EM 2022, se propuso a los estudiantes una serie de afirmaciones o ítems que buscaban medir las habilidades socioemocionales seleccionadas; por ende, estos ítems hacen referencia a aspectos característicos de las mismas.

1.2 Procesos de validación de los instrumentos de evaluación

Estos procesos se realizan para obtener evidencias de validez vinculadas al contenido de las pruebas. Este tipo de evidencias se vincula con la necesidad de garantizar que la prueba sea una muestra adecuada y representativa de los contenidos que pretende evaluar (McGartland et al., 2003; Muñiz, 1998). En esta misma línea, Suen (1990) indica que los ítems en un test o prueba, constituyen supuestamente una muestra representativa de todos los ítems que se pueden emplear para medir el constructo de interés. Si este supuesto se cumple, entonces se podrá decir que la prueba tiene adecuadas evidencias de validez referidas a su contenido. Esta se definiría, entonces, como la representatividad de los ítems empleados en un test para medir un constructo particular. A su vez, la representatividad de un test sería definido como la precisión con que se puede hacer inferencias acerca de la puntuación de la persona en el universo de indicadores a partir de la puntuación en la muestra de ítems que constituyen la prueba (Paz, 1996).

Muñiz (1998) señala que la práctica más usual para lograr este tipo de validez consiste en enumerar todas las áreas de contenido (o dimensiones) consideradas importantes o imprescindibles; luego, debe asegurarse de que la prueba contenga ítems referidos a cada una de ellas en una proporción adecuada. En ese sentido, el proceso de validación implica el mismo proceso de construcción, pues se debe construir y seguir matrices (o tablas) de especificaciones.

La elaboración de estas matrices de especificaciones tiene como base el CNEB. Con ello, se garantiza el alineamiento de las pruebas con el currículo vigente, lo cual es una condición sine qua non para que estas brinden evidencias suficientes del logro de la competencia a evaluar en diferentes niveles de desarrollo por parte de los estudiantes. Asimismo, las matrices de un año actual guardan una estrecha relación con las del año anterior en cuanto a su contenido pedagógico, así como respecto de las proporciones o pesos de cada dimensión de la competencia evaluada en la prueba para, junto con los procesos de equiparación, asegurar su comparabilidad en el tiempo. En tal sentido, los porcentajes de cada dimensión (competencias, capacidades, conocimientos y contextos) son muy similares año tras año. Por ejemplo, las matrices de especificaciones de la ENLA 2023 son muy similares a las de la Evaluación Muestral 2022. Esta correspondencia de las matrices es validada por los especialistas de las direcciones de línea del Ministerio de Educación.

En los siguientes acápite, se describe los procesos de validación de los instrumentos de evaluación elaborados sobre la base de las matrices de especificaciones.

1.2.1 Validación de expertos de la UMC

Después de que los equipos de especialistas de las áreas (Matemática, Comunicación, Ciudadanía y Ciencias Sociales, Ciencia y Tecnología y EIB) han construido los ítems, estos son revisados por expertos de la UMC. Para esto, se cuenta con guías que orientan el proceso, así como con fichas que permiten evaluar cada ítem en aspectos como calidad, vigencia y veracidad de la información según cada disciplina científica, la correspondencia con las matrices de especificaciones, la adecuación de la complejidad del ítem a la población evaluada y la construcción del enunciado y de las alternativas, tanto en lo formal como en su eficacia para la medición del constructo a evaluar. Asimismo, se revisa que no se presenten sesgos socioeconómicos, culturales ni de género en la construcción de los ítems. En el caso de las pruebas de Lectura correspondientes al área de Comunicación (2.º, 4.º y 6.º grados de primaria, 2.º grado de secundaria y 4.º grado de primaria EIB), se revisan también diversos aspectos referidos a los textos a partir de los cuales se construyen los ítems, tales como el tipo y el género textuales, la estructura textual, la densidad de la información, el vocabulario, la sintaxis, la familiaridad del tema, la extensión y el formato.

1.2.2 Validación de campo: aplicación piloto

Después de la revisión de expertos al interior de la UMC, los subequipos de evaluación responsables de las pruebas subsanan las observaciones e incorporan las sugerencias hechas por los revisores. Luego de esto, los ítems pasan a ser empleados en la construcción de las pruebas destinadas a la aplicación piloto, la cual consiste en la aplicación de pruebas a una muestra representativa de estudiantes a nivel nacional.

Un diseño estadístico permite determinar las instituciones educativas (IE) que formarán parte de la muestra de la aplicación piloto. De acuerdo con el procedimiento estadístico, las IE son elegidas aleatoriamente y corresponden a los distintos estratos (según las características determinadas en el diseño): rural y urbano, estatal y no estatal, polidocente y multigrado/unidocente. Asimismo, cabe señalar que los estudiantes de estas IE presentan similares condiciones a las de la población objetivo (edad, avance curricular, desarrollo de capacidades, etc.).

Además, las pruebas piloto presentan estructuras y características similares (orden de presentación de los ítems según su dificultad, extensión, dificultad, diagramación, tiempo de aplicación de la prueba) a las de las pruebas definitivas.

La aplicación piloto busca recoger información que permita asegurar que los ítems en particular y los instrumentos en su conjunto presentan adecuadas evidencias de validez referidas a cada uno de los constructos que se evalúan mediante cada una de las pruebas. Asimismo, se recogen evidencias acerca del funcionamiento de los ítems y su ajuste al modelo de medición empleado en la UMC.

Cabe señalar que las aplicaciones piloto, además de alimentar el banco de ítems de la UMC, permiten renovar un porcentaje de los ítems de las pruebas cada año y reemplazar aquellos ítems liberados; es decir, que son publicados en los informes de resultados que produce y distribuye la UMC.

1.2.3 Revisión de ítems y consulta con la población objetivo

En el caso de las HSE, en 2.º grado de secundaria se realizó, en diferentes momentos, un proceso de revisión de los ítems que conformaban las escalas seleccionadas para su medición. Las revisiones consistieron en reuniones virtuales con estudiantes, las cuales permitieron socializar los ítems y asegurar su comprensión y reformulación. En toma de decisiones, durante el 2021, y en resolución de problemas, durante el 2023, participaron estudiantes de diferentes zonas urbanas y rurales del país. Como resultado de estas consultas con estudiantes, se realizaron ajustes en los términos que no eran bien comprendidos.

1.2.4 Análisis de las propiedades psicométricas y de la distribución de los ítems en la escala

Sobre la base de los resultados de la aplicación piloto, se analizan las propiedades y otros aspectos psicométricos de las pruebas y los ítems. Además, se realiza una revisión pedagógica de la distribución de los ítems en la escala de dificultad. Entre los aspectos psicométricos, se analizan el ajuste de los ítems, la correlación ítem-medida, la confiabilidad de las medidas, la unidimensionalidad, la correspondencia entre la media de habilidad de los estudiantes y la media de dificultad de los ítems, y el funcionamiento diferencial de los ítems según los distintos estratos (urbano y rural, estatal y no estatal, hombre y mujer) y según su posición en la prueba. Igualmente, se ensayan hipótesis sobre las posibles causas por las que un ítem no se ajustó al modelo psicométrico, tuvo un funcionamiento diferencial o su dificultad no concordó con la complejidad propuesta. En el capítulo 4 de este documento, se detallan dichas propiedades psicométricas con la finalidad de ampliar más este aspecto.

En la revisión de aspectos pedagógicos, se analiza la correspondencia entre la complejidad pedagógica estimada de un ítem y la dificultad dada por el modelo psicométrico, el funcionamiento de los distractores, la pertinencia de incluir un determinado ítem en la prueba definitiva y las posibles causas por las que los estudiantes no contestaron correctamente un ítem. En el caso del área de Comunicación (Lectura), tanto en primaria como en secundaria, también se analizan los textos para encontrar posibles explicaciones a las respuestas no esperadas o a un error recurrente de los estudiantes.

Este proceso de análisis concluye con la selección de los ítems de mejor funcionamiento y ajuste al modelo psicométrico para la conformación de las pruebas de la aplicación definitiva.

1.2.5 Validación de expertos de otras oficinas del Minedu

De acuerdo con las recomendaciones técnicas, el proceso de validación de los ítems incluye que un conjunto de expertos en cada una de las áreas evaluadas emita su juicio sobre tres aspectos: la construcción, la correspondencia con el indicador señalado y la dificultad de la tarea solicitada en cada ítem empleado para medir el constructo implicado (criterio de jueces). Para ello, los ítems seleccionados a partir de la aplicación piloto y el análisis posterior son organizados en formas para que sean validados por expertos de las distintas direcciones involucradas por parte del Minedu, tales como la Dirección de Educación Primaria (DEP), la Dirección de Educación Secundaria (DES) y la Dirección de Educación Intercultural Bilingüe (DEIB) de la Dirección General de Educación Básica Alternativa, Intercultural Bilingüe y de Servicios Educativos (Digeibira). En este proceso de validación, también participan especialistas de la Dirección de Evaluación Docente (DIED) por su experiencia en el diseño y la construcción de ítems, así como por su aporte relacionado con dicha experiencia.

En esta validación, se solicita a los jueces que evalúen los aspectos mencionados en el párrafo anterior mediante una ficha técnica. Esta cuenta con cuatro secciones: tres en las que se presentan los indicadores que evalúan tanto las características generales del ítem como las características específicas del enunciado y de las alternativas, y una sección en que el juez escribe sugerencias puntuales sobre el ítem.

Las fichas presentan ciertas particularidades de acuerdo con las características de la competencia evaluada. Por ejemplo, en el caso de Lectura (área de Comunicación), se incluyen indicadores referidos a cada uno de los textos y no solamente a los ítems. Cada experto completa la ficha y formula las observaciones que, de acuerdo con su juicio particular, estime convenientes. Luego, los subequipos de cada área se encargan de sistematizar las fichas, absolver los cuestionamientos e incorporar las sugerencias. Ello implica, en algunos casos, realizar algún tipo de ajuste al ítem y/o al texto.

1.2.6 Adaptación de instrumentos para estudiantes con discapacidad

Desde el 2015, la UMC adapta (es decir, realiza ajustes razonables) los instrumentos de evaluación de las distintas competencias evaluadas, así como los procedimientos de aplicación de las pruebas. Estas adaptaciones se realizan con el fin de asegurar la participación plena y autónoma en las evaluaciones nacionales de logros de aprendizaje de aquellos estudiantes con discapacidad sensorial (baja visión, ceguera, hipoacusia y sordera), discapacidad física y trastorno del espectro autista.

Dependiendo de su tipo y su profundidad, Browning (2003) clasifica las adaptaciones en dos categorías: acomodaciones y modificaciones. Las acomodaciones son cambios introducidos a los instrumentos de evaluación o a los procedimientos de aplicación que posibilitan a los estudiantes con discapacidad acceder autónomamente a las evaluaciones, sin alterar el constructo medido. En contraste, las modificaciones son cambios que alteran el estándar o las expectativas de logro y, por tanto, el constructo medido, ya que no se evalúan las mismas habilidades que forman parte de la competencia a diferencia de los demás estudiantes.

En la ENLA 2023, se implementaron las siguientes acomodaciones en las pruebas y procesos de aplicación:

Tabla 1.2 Acomodaciones a las pruebas de la ENLA 2023

Tipo de discapacidad		Tipo de acomodación	Descripción
Visual	Baja visión	Presentación	Prueba en macrotipo
		Equipos y materiales	Plumón delgado, lupa, lápiz jumbo (extragrueso)

Tipo de discapacidad	Tipo de acomodación	Descripción
Baja visión	Extensión de la prueba	Menor cantidad de preguntas a resolver en el mismo tiempo establecido para todos los estudiantes
	Ubicación	Carpeta junto a acceso de luz natural
	Apoyo	Aplicador personalizado
Visual	Presentación	Prueba en sistema braille e imágenes en relieve Prueba asistida (impresa en tinta e imágenes en relieve, para estudiantes que no dominan el sistema braille) con el apoyo en la lectura de un aplicador
	Equipos y materiales	Regleta, punzón, ábaco, cubaritmo
Ceguera total	Extensión de la prueba	Menor cantidad de preguntas a resolver en el mismo tiempo establecido para todos los estudiantes
	Ubicación	Carpeta cerca a puerta de acceso al aula, para el caso de los estudiantes que rinden la prueba en braille Aula diferente para los estudiantes que recibieron la prueba asistida
	Apoyo	Aplicador personalizado
Auditiva	Presentación	Prueba con léxico y sintaxis simplificados
	Equipos/materiales	Lápiz, borrador
	Extensión de la prueba	Menor cantidad de preguntas a resolver en el mismo tiempo establecido para todos los estudiantes
	Ubicación	Carpeta al fondo del aula
	Apoyo	Intérprete de lengua de señas peruana (si el estudiante lo requería)
Sordoceguera	Según la severidad de esta discapacidad, se aplicaron las mismas acomodaciones que en los casos de baja visión o ceguera.	

Tipo de discapacidad	Tipo de acomodación	Descripción
Física	Presentación	Prueba sin adaptaciones en el formato o prueba en macrotipo (si el estudiante lo requería debido a la distancia con el objeto de lectura)
	Equipos/materiales	Lápiz jumbo, soporte de lápiz (si el estudiante lo requería)
	Extensión de la prueba	Igual cantidad de preguntas a resolver que todos los estudiantes y en el mismo tiempo establecido para todos los estudiantes
	Ubicación	Carpeta cerca a puerta de acceso al aula
	Apoyo	Aplicador personalizado para asistirlo en el marcado de las respuestas (si el estudiante lo requería)
Trastorno del espectro autista	Presentación	Prueba sin adaptaciones en el formato
	Equipos/materiales	Tarjetas de comunicación alternativa (si el estudiante lo requería)
	Extensión de la prueba	Igual cantidad de preguntas a resolver que todos los estudiantes y en el mismo tiempo establecido para todos los estudiantes
	Ubicación	Lugar y carpeta que usa habitualmente
	Apoyo	Aplicador personalizado
Autismo	Presentación	Prueba con léxico y sintaxis simplificados
	Equipos/materiales	Tarjetas de comunicación alternativa
	Extensión de la prueba	Menor cantidad de preguntas a resolver en el mismo tiempo establecido para todos los estudiantes
	Ubicación	Lugar y carpeta que usa habitualmente
	Apoyo	Aplicador personalizado

Tabla 1.3 Modificaciones a las pruebas de la ENLA 2023

Discapacidad	Tipo de modificación	Modificación
Intelectual	Presentación	Prueba diferenciada para estudiantes lectores y lectores iniciales Prueba de complejidad escalonada
	Equipos/materiales	Lápiz, borrador
	Extensión de la prueba	Menor cantidad de preguntas a resolver en el mismo tiempo establecido para todos los estudiantes
	Ubicación	Carpeta al fondo del aula
	Apoyo	Aplicador personalizado para asistirlo en la lectura y/o marcado de las respuestas (si el estudiante lo requería)

1.2.7 Aplicación definitiva

La aplicación definitiva de la ENLA 2023 se realizó en el último trimestre del año (octubre, noviembre y diciembre). Se llevó a cabo durante dos días en 2.º y 4.º grados de primaria, un día en 4.º grado de primaria EIB y tres días en 2.º grado de secundaria. En el primer día, tanto en 2.º y 4.º grados de primaria como en 2.º grado de secundaria, se aplicaron, primero, las pruebas de Lectura y, luego, las de Matemática. En el segundo día, se invirtió el orden de las pruebas aplicadas. Este cambio obedece a la necesidad de neutralizar los posibles efectos del cansancio de los estudiantes, pues rinden dos pruebas en un día. En el caso particular de 4.º grado de primaria EIB, la aplicación duró un día porque solo se aplicó la prueba de Lectura en castellano como segunda lengua. Por su parte, en el caso de 2.º grado de secundaria, el tercer día se destinó a la aplicación de la prueba de Ciencias Sociales.

Cabe resaltar que, a partir de 2015, los instrumentos utilizados en la aplicación definitiva de las evaluaciones nacionales de logros de aprendizaje, incluidos los cuadernillos de las pruebas, retornan en su totalidad al Minedu y se mantienen bajo estricta confidencialidad. Este procedimiento, aunque complicado, se realiza porque permite alimentar un banco de ítems creciente y asegurar el proceso de equiparación de las pruebas por ítems en común y, con ello, la comparabilidad de los resultados en el tiempo.

En resumen, mediante los procesos de construcción y validación se logró que los instrumentos de evaluación de la ENLA 2023 cumplan con los requerimientos técnicos y consideraciones de contenido, los cuales permiten generar información válida y confiable sobre los niveles de logro de los estudiantes evaluados. Asimismo, se garantizó que los instrumentos guardaran correspondencia con los diversos

documentos curriculares del sistema educativo nacional y que los resultados derivados de su aplicación sean comparables a lo largo del tiempo.

1.3 Ensamblaje de las pruebas mediante el diseño bloques

Conforme avanzan en su escolaridad, los estudiantes deben lograr aprendizajes cada vez más complejos. Por ejemplo: en Lectura, el lenguaje de los textos se complejiza y se especializa progresivamente, los temas tratados en los textos son más diversos, y las capacidades de inferencia y reflexión se diversifican y manifiestan en desempeños cada vez más complejos. Por su parte, en Matemática, las nociones de número, variación y espacio, en un inicio ligadas a situaciones más particulares y objetos concretos, demandan mayor formalización, abstracción y aprendizaje de conceptos nuevos, necesarios para seguir construyendo el pensamiento matemático. Por esta razón, es fundamental proponer una evaluación que cubra un amplio espectro de capacidades y conocimientos, de manera que los resultados recojan de forma precisa lo que realmente saben y pueden hacer los estudiantes.

Para lograr este objetivo, se ha utilizado el diseño de bloques no balanceados incompletos que permitió generar diversas formas de la misma prueba. Las evidencias generadas mediante las aplicaciones piloto, así como por las simulaciones realizadas (las cuales consideraron distintos escenarios) han permitido tomar esta decisión con la certeza de contar con buenos o aceptables indicadores psicométricos de confiabilidad de los resultados, así como de precisión y consistencia en la clasificación de los estudiantes en múltiples iteraciones. De la misma manera, estos diseños de pruebas han posibilitado que las formas se adapten al nivel de habilidad de grupos de estudiantes diferentes. En este sentido, se han elaborado formas con un mayor nivel de dificultad o para zona A, las cuales han sido aplicadas con estudiantes de escuelas urbanas, así como formas con menor nivel de dificultad o para zona B, que han sido aplicadas por estudiantes de escuelas rurales y por todos los estudiantes de las regiones Loreto y Ucayali.

En concordancia con lo anterior, la organización y ensamblaje de estas formas contempló la presencia de bloques de ítems comunes (generalmente, en una posición fija), que formaran parte de las formas de mayor y menor dificultad, y de bloques no comunes (generalmente en diferentes posiciones), que formaran parte solo de las formas de mayor dificultad o solo de aquellas de menor dificultad. La combinación de bloques comunes y no comunes permitió el ensamblaje de formas que se distribuyeron aleatoriamente entre los estudiantes de cada grupo antes mencionado. De este modo y gracias al uso de técnicas de equiparación del modelo Rasch, las distintas pruebas estuvieron conformadas por una mayor cantidad de ítems en la misma métrica, sin que ello implicara incrementar la cantidad de ítems a los que cada estudiante se enfrentó. Además, estas formas tenían un nivel de dificultad ajustado al nivel de habilidad del grupo destinatario.

En relación con la conformación de los bloques, es preciso señalar que el bloque común (distribuido en pequeños bloques comunes) tenía también la finalidad de optimizar la equiparación de las medidas, y que fue construido de tal forma que los ítems reflejaran la tabla de especificaciones, así como la dispersión de las medidas a lo largo de la escala de dificultad.

Por último, en el caso de las pruebas con adaptaciones destinadas a estudiantes con discapacidad, se ha considerado el diseño de prueba única.

1.3.1 Diseño de bloques de la ENLA 2023

En las siguientes tablas, se presenta el diseño de bloques de las pruebas aplicadas. Las tablas de especificaciones de estos instrumentos pueden ser consultadas en los anexos.

Diseño de bloques de las pruebas de Lectura de 2.º grado de primaria

La prueba de Lectura de 2.º grado de primaria elaborada para la ENLA 2023 estuvo conformada por 90 ítems, de los cuales 84 se construyeron a partir de 20 textos de diferente tipo, género, formato y complejidad, mientras que seis ítems fueron de lectura inicial, por lo que no están vinculados a ningún texto. Los 20 textos y 90 ítems se organizaron en 26 bloques. De ellos, 14 bloques son comunes y se encuentran en posición fija (seis comunes a todas las formas y ocho comunes a las formas correspondientes a cada día de aplicación; cuatro por día) y doce bloques son no comunes y se encuentran distribuidos entre las distintas formas. Las diferentes combinaciones de los bloques permitieron ensamblar cuatro formas para escuelas con estudiantes de mayor habilidad o zona A, y cuatro para escuelas con estudiantes de menor habilidad o zona B. De acuerdo con esta distribución, cada forma y el conjunto de estas reflejan los contenidos (capacidades, tipos de textos, indicadores, entre otros) de las tablas de especificaciones de la prueba. Dentro de su tipo (de zona A o de zona B), estas formas presentan una dificultad similar entre sí, pero diferenciada según el grupo de estudiantes al que están dirigidas.

Tabla 1.4 Matriz de bloques de la prueba de Lectura de 2.º grado de primaria. Día 1

Forma	Bloque	Ítems por forma
1	B03	25
	B05	
	B06	
	B07	
	B08	
	B10	

Forma	Bloque	Ítems por forma
2	B04	25
	B05	
	B06	
	B07	
	B09	
	B11	
3	B01	25
	B03	
	B05	
	B06	
	B07	
	B08	
4	B12	25
	B02	
	B04	
	B05	
	B06	
	B07	
4	B09	25
	B13	

Tabla 1.5 Matriz de bloques de la prueba de Lectura de 2.º grado de primaria. Día 2

Forma	Bloque	Ítems por forma
5	B16	25
	B18	
	B19	
	B20	
	B21	
	B23	
6	B17	25
	B18	
	B19	
	B20	
	B22	
6	B24	25

Forma	Bloque	Ítems por forma
7	B14	25
	B16	
	B18	
	B19	
	B20	
	B21	
	B25	
8	B15	25
	B17	
	B18	
	B19	
	B20	
	B22	
	B26	

De las ocho formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 90 ítems que tiene esta prueba, cada estudiante respondió 50.

Diseño de bloques de las pruebas de Lectura de 4.º grado de primaria

La prueba de Lectura de 4.º grado de primaria elaborada para la ENLA 2023 estuvo conformada por 17 textos de diferentes tipo, género, formato y complejidad, a partir de los cuales se elaboraron 96 ítems. Dichos textos e ítems se distribuyeron en 16 bloques. De ellos, 14 bloques son comunes y se encuentran en posición fija (seis comunes a todas las formas y ocho comunes a las formas correspondientes a cada día de aplicación; cuatro por día) y doce bloques son no comunes y se encuentran distribuidos entre las distintas formas. De ellos, cuatro bloques son comunes fijos (dos por día) y doce bloques son no comunes (seis por día) que se encuentran distribuidos entre las distintas formas. Las diferentes combinaciones de los bloques permitieron ensamblar seis formas para escuelas con estudiantes de mayor habilidad o zona A y seis para escuelas con estudiantes de menor habilidad o zona B. De acuerdo con esta distribución, cada forma y el conjunto de estas reflejan los contenidos (capacidades, tipos de textos, indicadores, entre otros) de las tablas de especificaciones de la prueba. Dentro de su tipo (de zona A o de zona B), estas formas presentan una dificultad similar entre sí, pero diferenciada según el grupo de estudiantes al que están dirigidas.

Tabla 1.6 Matriz de bloques de la prueba de Lectura de 4.º grado de primaria. Día 1

Forma	Bloque	Ítems por forma
1	B05	24
	B01	
	B02	
	B06	
2	B06	24
	B01	
	B02	
	B07	
3	B07	24
	B01	
	B02	
	B08	
4	B05	24
	B01	
	B02	
	B09	
5	B09	24
	B01	
	B02	
	B10	
6	B10	24
	B01	
	B02	
	B08	

Tabla 1.7 Matriz de bloques de la prueba de Lectura de 4.º grado de primaria. Día 2

Forma	Bloque	Ítems por forma
7	B11	24
	B03	
	B04	
	B12	

Forma	Bloque	Ítems por forma
8	B12	24
	B03	
	B04	
	B13	
9	B13	24
	B03	
	B04	
	B14	
10	B14	24
	B03	
	B04	
	B15	
11	B15	24
	B03	
	B04	
	B16	
12	B16	24
	B03	
	B04	
	B11	

De las doce formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por día de aplicación). En consecuencia, de los 96 ítems que tiene esta prueba, cada estudiante respondió 48.

Diseño de bloques de la prueba de Lectura en castellano como segunda lengua de 4.º grado de primaria EIB

La prueba de Lectura en castellano como segunda lengua de 4.º grado de primaria EIB elaborada para la ENLA 2023 estuvo conformada por 23 textos de diferente tipo, género, formato y complejidad, a partir de los cuales se elaboraron 144 ítems. Dichos textos e ítems se distribuyeron en 23 bloques. De ellos, siete bloques son comunes y 16 son no comunes. Las diferentes combinaciones de los bloques permitieron ensamblar ocho formas, cuatro de mayor dificultad (para escuelas de lengua andina) y cuatro de menor dificultad (para escuelas de lengua amazónica). De acuerdo con esta distribución, cada forma y el conjunto de estas reflejan los contenidos (capacidades, tipos de textos, indicadores, entre otros) de las tablas de especificaciones de la prueba. Dentro de su tipo (de mayor o menor dificultad), estas formas presentan una dificultad similar entre sí, pero diferenciada según el grupo de estudiantes al que están dirigidas.

Tabla 1.8 Matriz de bloques de la prueba de Lectura en castellano como segunda lengua de 4.º grado de primaria EIB. Día único

Forma	Bloque	Ítems por forma
1	B01	24
	B05	
	B09	
	B13	
	B11	
2	B02	24
	B06	
	B10	
	B14	
	B12	
3	B03	24
	B07	
	B11	
	B14	
	B10	
4	B04	24
	B08	
	B12	
	B15	
	B09	
5	B01	24
	B05	
	B20	
	B22	
	B21	
6	B16	24
	B18	
	B10	
	B14	
	B12	
7	B17	24
	B19	
	B21	
	B14	
	B10	
8	B04	24
	B08	
	B12	
	B23	
	B20	

De las ocho formas ensambladas, cada estudiante evaluado se enfrentó solo a una de ellas. En consecuencia, de los 144 ítems que tiene esta prueba, cada estudiante respondió 30.

Diseño de bloques de las pruebas de Lectura de 2.º grado de secundaria

La prueba de Lectura de 2.º grado de secundaria elaborada para la ENLA 2023 estuvo conformada por 20 textos de diferente tipo, género, formato y complejidad, a partir de los cuales se elaboraron 122 ítems. Dichos textos e ítems se distribuyeron en 20 bloques. De ellos, cuatro bloques son comunes y se encuentran en posición fija (dos por día), y 16 bloques son no comunes (ocho por día) que se encuentran distribuidos entre las distintas formas. Las diferentes combinaciones de los bloques permitieron ensamblar 18 formas, diez para escuelas con estudiantes de mayor habilidad o zona A y ocho para escuelas con estudiantes de menor habilidad o zona B. De acuerdo con esta distribución, cada forma y el conjunto de estas reflejan los contenidos (capacidades, tipos de textos, indicadores, entre otros) de las tablas de especificaciones de la prueba. Dentro de su tipo (de mayor o menor dificultad), estas formas presentan una dificultad similar entre sí, pero diferenciada según el grupo de estudiantes al que están dirigidas.

Tabla 1.9 Matriz de bloques de la prueba de Lectura de 2.º grado de secundaria. Día 1

Forma	Bloque	Ítems por forma
1	B05	25
	B01	
	B02	
	B06	
2	B06	25
	B01	
	B02	
3	B07	25
	B01	
	B02	
4	B08	25
	B08	
	B01	
	B02	
	B09	

Forma	Bloque	Ítems por forma
5	B07	25
	B01	
	B02	
	B10	
6	B05	25
	B01	
	B02	
7	B11	25
	B01	
	B02	
	B12	
8	B12	25
	B01	
	B02	
9	B10	25
	B01	
	B02	
	B09	

**Tabla 1.10 Matriz de bloques de la prueba de Lectura de 2.º grado de secundaria.
Día 2**

Forma	Bloque	Ítems por forma
10	B13	25
	B03	
	B04	
	B14	
11	B14	25
	B03	
	B04	
12	B15	25
	B15	
	B03	
	B04	
	B16	

Forma	Bloque	Ítems por forma
13	B16	25
	B03	
	B04	
	B17	
14	B17	25
	B03	
	B04	
	B13	
15	B13	25
	B03	
	B04	
	B18	
16	B18	25
	B03	
	B04	
	B19	
17	B19	25
	B03	
	B04	
	B20	
18	B20	25
	B03	
	B04	
	B17	

De las 18 formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por día de aplicación). En consecuencia, de los 122 ítems que tiene esta prueba, cada estudiante respondió 50.

Diseño de bloques de las pruebas de Matemática de 2.º grado de primaria

La prueba de Matemática de 2.º grado de primaria elaborada para la ENLA 2023 estuvo conformada por un total de 92 ítems distribuidos en ocho formas. Cada forma contiene 23 ítems: 11 ítems comunes organizados en dos bloques comunes en posición fija y doce ítems no comunes organizados en dos bloques rotados que varían de una forma a otra. Las diferentes combinaciones de los bloques permitieron ensamblar cuatro formas para escuelas con estudiantes de mayor habilidad o zona A y cuatro para escuelas con estudiantes de menor habilidad o zona B. De acuerdo con esta distribución, cada forma y el conjunto de estas reflejan los

contenidos (capacidades, conocimientos, indicadores, entre otros) de las tablas de especificaciones de la prueba. Dentro de su tipo (de zona A o de zona B), estas formas presentan una dificultad similar entre sí, pero diferenciada según el grupo de estudiantes al que están dirigidas.

Tabla 1.11 Matriz de bloques de la prueba de Matemática de 2.º grado de primaria. Día 1

Forma	Bloque	Ítems por forma
1	B01	23
	B09	
	B14	
2	B02	23
	B10	
	B13	
3	B01	23
	B05	
	B09	
4	B13	23
	B06	
	B06	
	B10	
5	B14	23
	B05	
	B03	

Tabla 1.12 Matriz de bloques de la prueba de Matemática de 2.º grado de primaria. Día 2

Forma	Bloque	Ítems por forma
5	B03	23
	B11	
	B16	
6	B04	23
	B04	
	B12	
7	B15	23
	B03	

Forma	Bloque	Ítems por forma
7	B07	23
	B11	
	B15	
	B08	
8	B08	23
	B12	
	B16	
	B07	

De las ocho formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 92 ítems que tiene esta prueba, cada estudiante respondió 46.

Diseño de bloques de las pruebas de Matemática de 4.º grado de primaria

La prueba de Matemática de 4.º grado de primaria elaborada para la ENLA 2023 estuvo conformada por un total de 102 ítems organizados en 20 bloques: 14 bloques comunes y seis no comunes. Con estos 20 bloques, se ensamblaron doce formas. Cada forma contiene 25 ítems. Las formas 1, 2, 7 y 8 están constituidas por 11 ítems comunes organizados en cuatro bloques comunes en posición fija y 14 ítems no comunes organizados en dos bloques que varían de posición de una forma a otra. Las formas 3, 4, 5, 6, 9, 10, 11 y 12 están constituidas por 18 ítems comunes organizados en cinco bloques comunes, cuatro en posición fija y uno que varía de posición de una forma a otra, además de siete ítems no comunes organizados en un bloque que varía de posición de una forma a otra. Las diferentes combinaciones de los bloques permitieron ensamblar seis formas para escuelas con estudiantes de mayor habilidad o zona A y seis para escuelas con estudiantes de menor habilidad o zona B. De acuerdo con esta distribución, cada forma y el conjunto de estas reflejan los contenidos (competencias, capacidades, conocimientos, indicadores, entre otros) de las tablas de especificaciones de la prueba. Dentro de su tipo (de zona A o de zona B), estas formas presentan una dificultad similar entre sí, pero diferenciada según el grupo de estudiantes al que están dirigidas.

**Tabla 1.13 Matriz de bloques de la prueba de Matemática de 4.º grado de primaria.
Día 1**

Forma	Bloque	Ítems por forma
1	B01	25
	B09	
	B13	
	B19	
	B10	
	B02	
2	B02	25
	B10	
	B14	
	B19	
	B09	
3	B01	25
	B06	
	B10	
	B15	
	B19	
4	B09	25
	B02	
	B05	
	B09	
	B13	
	B19	
5	B10	25
	B14	
	B19	
	B09	
	B05	
6	B05	25
	B10	
	B15	
	B19	
	B09	
	B06	

**Tabla 1.14 Matriz de bloques de la prueba de Matemática de 4.º grado de primaria.
Día 2**

Forma	Bloque	Ítems por forma
7	B03	25
	B11	
	B16	
	B20	
	B12	
8	B04	25
	B04	
	B12	
	B17	
	B20	
9	B11	25
	B03	
	B08	
	B12	
	B18	
10	B20	25
	B11	
	B04	
	B07	
	B11	
11	B16	25
	B20	
	B12	
	B08	
	B08	
12	B12	25
	B17	
	B20	
	B11	
	B07	
12	B07	25
	B12	
	B18	
	B20	
	B11	
	B08	

De las doce formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 102 ítems que tiene esta prueba, cada estudiante respondió 50.

Diseño de bloques de las pruebas de Matemática de 2.º grado de secundaria

La prueba de Matemática de 2.º grado de secundaria elaborada para la ENLA 2023 estuvo conformada por un total de 122 ítems organizados en 22 bloques: 16 bloques comunes y seis no comunes. Con estos 22 bloques, se ensamblaron 18 formas. Cada forma contiene 25 ítems, 19 ítems comunes organizados en cinco bloques comunes (cuatro en posición fija y uno que varía de posición de una forma a otra); y seis ítems no comunes organizados en un bloque que varía de posición de una forma a otra. Las diferentes combinaciones de los bloques permitieron ensamblar diez formas para escuelas con estudiantes de mayor habilidad o zona A, y ocho para escuelas con estudiantes de menor habilidad o zona B. De acuerdo con esta distribución, cada forma y el conjunto de estas reflejan los contenidos (competencias, capacidades, conocimientos, indicadores, entre otros) de las tablas de especificaciones de la prueba. Dentro de su tipo (de zona A o de zona B), estas formas presentan una dificultad similar entre sí, pero diferenciada según el grupo de estudiantes al que están dirigidas.

Tabla 1.15 Matriz de bloques de la prueba de Matemática de 2.º grado de secundaria. Día 1

Forma	Bloque	Ítems por forma
1	B06	25
	B09	
	B13	
	B21	
	B10	
	B01	
2	B02	25
	B10	
	B14	
	B21	
	B09	
	B06	

Forma	Bloque	Ítems por forma
3	B06	25
	B10	
	B15	
	B21	
	B09	
	B02	
4	B01	25
	B09	
	B16	
	B21	
	B10	
	B06	
5	B06	25
	B10	
	B13	
	B21	
	B09	
	B02	
6	B05	25
	B09	
	B13	
	B21	
	B10	
	B06	
7	B06	25
	B10	
	B14	
	B21	
	B09	
	B05	
8	B05	25
	B10	
	B15	
	B21	
	B09	
	B06	

Forma	Bloque	Ítems por forma
9	B06	25
	B09	
	B16	
	B21	
	B10	
	B05	

Tabla 1.16 Matriz de bloques de la prueba de Matemática de 2.º grado de secundaria. Día 2

Forma	Bloque	Ítems por forma
10	B08	25
	B11	
	B17	
	B22	
	B12	
	B03	
11	B04	25
	B12	
	B18	
	B22	
	B11	
12	B08	25
	B12	
	B19	
	B22	
	B11	
	B04	
13	B03	25
	B11	
	B20	
	B22	
	B12	
	B08	

Forma	Bloque	Ítems por forma
14	B08	25
	B12	
	B17	
	B22	
	B11	
	B04	
15	B07	25
	B11	
	B17	
	B22	
	B12	
	B08	
16	B08	25
	B12	
	B18	
	B22	
	B11	
	B07	
17	B07	25
	B12	
	B19	
	B22	
	B11	
	B08	
18	B08	25
	B11	
	B20	
	B22	
	B12	
	B07	

De las 18 formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 122 ítems que tiene esta prueba, cada estudiante respondió 50.

Diseño de bloques de las pruebas de Ciencias Sociales de 2.º grado de secundaria

La prueba de Ciencias Sociales de 2.º grado de secundaria elaborada para la ENLA 2023 estuvo conformada por un total de 126 ítems distribuidos en siete bloques comunes (de seis ítems cada uno) y 14 no comunes (de seis ítems cada uno). Las diferentes combinaciones de los bloques permitieron ensamblar seis formas para escuelas con estudiantes de mayor habilidad o zona A, y seis para escuelas con estudiantes de menor habilidad o zona B. De acuerdo con esta distribución, cada forma y el conjunto de estas reflejan los contenidos (competencias, capacidades, conocimientos, indicadores, entre otros) de las tablas de especificaciones de la prueba. Dentro de su tipo (de mayor o menor dificultad), estas formas presentan una dificultad similar entre sí, pero diferenciada según el grupo de estudiantes al que están dirigidas.

Tabla 1.17 Matriz de bloques de la prueba de Ciencias Sociales de 2.º grado de secundaria. Día 3

Forma	Bloque	Ítems por forma
19	B10	30
	B07	
	B01	
	B08	
	B11	
20	B12	30
	B08	
	B02	
	B07	
	B13	
21	B14	30
	B07	
	B03	
	B08	
	B15	

Forma	Bloque	Ítems por forma
22	B11	30
	B08	
	B04	
	B07	
	B12	
23	B13	30
	B07	
	B05	
	B08	
	B14	
24	B15	30
	B08	
	B06	
	B07	
	B10	
25	B16	30
	B07	
	B01	
	B09	
	B17	
26	B18	30
	B09	
	B02	
	B07	
	B19	
27	B20	30
	B07	
	B03	
	B09	
	B21	
28	B17	30
	B09	
	B04	
	B07	
	B18	

Forma	Bloque	Ítems por forma
29	B19	30
	B07	
	B05	
	B09	
	B20	
30	B21	30
	B09	
	B06	
	B07	
	B16	

De las doce formas ensambladas, cada estudiante evaluado se enfrentó a una de ellas. En consecuencia, de los 126 ítems que tiene esta prueba, cada estudiante respondió 30.

Diseño de bloques de los cuestionarios de Habilidades Socioemocionales de 2.º grado de secundaria

El cuestionario de Habilidades Socioemocionales de 2.º grado de secundaria elaborado para la ENLA 2023 cuenta con un total de 32 ítems distribuidos en dos bloques. Estos se distribuyen de la siguiente manera: B01 Toma de decisiones¹, doce ítems, y B02 Resolución de problemas, 20 ítems².

Tabla 1.18 Matriz de bloques del cuestionario de Habilidades Socioemocionales de 2.º grado de secundaria (día 2).

Forma	Bloque	Ítems por forma
1	B01 - Toma de decisiones	32
	B02 - Resolución de problemas	

Cada estudiante evaluado se enfrentó a los dos bloques de preguntas. En consecuencia, cada estudiante respondió todas las preguntas planteadas.

¹ La escala de Toma de decisiones tiene dos dimensiones: una de ellas busca indagar por estrategias adecuadas para la toma de decisiones (estilo vigilante), mientras que la otra busca explorar el uso de estrategias inadecuadas para la toma de decisiones (otros estilos). Así, si bien en el análisis se aplicó un modelo multidimensional para la estimación de los puntajes de los estudiantes, solo se reportan los resultados de los estudiantes que usan el estilo vigilante, pues estos ítems muestran directamente una habilidad socioemocional.

² La escala de resolución de problemas tiene un total de 20 ítems. Sin embargo, luego de realizar el análisis psicométrico, se eliminaron tres ítems debido a que no se ajustaron al modelo Rasch. Por este motivo, en el Capítulo 4 se presentan los resultados de los 17 ítems que se mantuvieron en el modelo final.

Población a evaluar

Capítulo 2

Capítulo 2

Población a evaluar

2.1 Población objetivo

La ENLA 2023 tuvo como población objetivo a los estudiantes que asistieron a clases en 2.º grado de primaria, 4.º grado de primaria, 4.º grado de primaria EIB y 2.º grado de secundaria de instituciones educativas (IE) de Educación Básica Regular (EBR).

Se evaluó a estos estudiantes porque en sus grados se concluyen los ciclos tercero, cuarto y sexto de la EBR, respectivamente, lo cual permite planificar intervenciones en el sistema para mejorar los grados y ciclos posteriores, en que el conocimiento se especializa y se hace más complejo.

Se excluyó a los estudiantes que asisten a instituciones educativas que atienden a menos de cinco estudiantes matriculados¹ en el grado a ser evaluado. Las razones se encuentran detalladas en el Marco de trabajo de la ECE (Ministerio de Educación, 2009). A continuación, se presentan los porcentajes de estudiantes excluidos a nivel nacional:

Tabla 2.1 Porcentaje de estudiantes excluidos según grado.

Grado	Porcentaje de exclusión
2.º grado de primaria	4,1
4.º grado de primaria	4,0
4.º grado de primaria EIB	20,5
2.º grado de secundaria	0,3

2.2 Marco censal y marco muestral

En la ENLA 2023 se llevaron a cabo evaluaciones muestrales en 2.º grado de primaria, 4.º grado de primaria y 2.º grado de secundaria, mientras que en 4.º grado de primaria EIB se llevó a cabo una evaluación censal. En este sentido, el listado conformado por todos los elementos que potencialmente pueden ser evaluados será denominado marco muestral cuando nos refiramos a las evaluaciones muestrales, y marco censal cuando nos refiramos a la evaluación censal de EIB. Ambos marcos se

¹ La cantidad de estudiantes matriculados tiene como fuente el dato registrado en el Sistema de Información de Apoyo a la Gestión de la Institución Educativa (Siagie). En caso de no contar con esta información, el dato proviene del Censo Educativo 2023.

construyen siguiendo la misma metodología; la única diferencia es que, en el caso del marco muestral, se seleccionará una muestra probabilística a partir del mismo, mientras que, en el marco censal, participarán todos sus elementos.

Los marcos para cada grado han sido elaborados con información proveniente de diferentes fuentes: el Padrón de Instituciones Educativas de la Unidad de Estadística Educativa (UEE), la cantidad de estudiantes y secciones del Censo Educativo 2022, y la información del Sistema de Información de apoyo a la Gestión de la Institución Educativa (Siagie) del 2022. Anualmente, la UEE del Minedu elabora y difunde el Padrón de Instituciones Educativas con fines estadísticos, el cual es actualizado constantemente. La información sobre las IE tiene como fuente operaciones estadísticas que se aplican a nivel nacional y los procedimientos asociados a estas. Es importante precisar que la información de la cantidad de estudiantes y secciones que proviene del Censo Educativo es declarada por el Director de la IE.

Para elaborar los marcos muestrales o censales en la ENLA 2023, se aplicó el siguiente procedimiento:

1. Se utilizó el Padrón de Instituciones Educativas para determinar el listado de IE activas². Se seleccionaron las IE de primaria y secundaria de EBR. En el caso de 2.º grado de primaria, 4.º grado de primaria y 2.º grado de secundaria, se excluyeron a aquellas IE que pertenecieran al tipo de atención fortalecimiento en Educación Intercultural Bilingüe (EIB) según el Registro Nacional de Instituciones Educativas que brindan el Servicio de Educación Intercultural Bilingüe (RNIIEE EIB). Por otro lado, para el caso de 4.º grado de primaria EIB se seleccionaron a aquellas IE que pertenecieran al tipo de atención fortalecimiento en EIB.
2. Se calculó la cantidad de secciones y estudiantes programados. Esta información corresponde al valor registrado en el Siagie en 2022. Si una IE no contaba con información del SIAGIE, se tomaba el valor del Censo Educativo 2022.
3. Se seleccionaron las IE que tuvieron cinco o más estudiantes asignados según el punto 2.

2.3 Diseño muestral

El diseño muestral no aplica para la evaluación en 4.º grado de primaria EIB, ya que esta es del tipo censal. En este sentido, esta sección solo es aplicable para 2.º grado de primaria, 4.º grado de primaria y 2.º grado de secundaria.

² El Padrón de Instituciones Educativas y el Censo Educativo se encuentran disponibles en la web ESCALE de la Unidad de Estadística Educativa: <http://escale.minedu.gob.pe/uee/>

El diseño de muestreo empleado fue del tipo probabilístico, estratificado, por conglomerados y bi-etápico, en el cual la unidad primaria de muestreo corresponde a las IE y la unidad secundaria corresponde a las secciones completas dentro de cada IE. En la primera etapa, la selección de escuelas fue proporcional al tamaño, por lo cual las escuelas con mayor cantidad de estudiantes en el grado a evaluar tuvieron una mayor probabilidad de ser seleccionadas. Según Lohr (2021), un diseño de conglomerado con probabilidades iguales puede resultar en estimaciones para la media con una varianza mayor que para un diseño proporcional al tamaño; además, en el contexto peruano se tiene una gran variabilidad de tamaños de escuelas por lo que, al utilizar un muestreo de conglomerados con probabilidades iguales, se podría derivar en la selección de muchas escuelas pequeñas, lo cual podría complicar la operación de campo. Finalmente, en la segunda etapa se realizó una selección aleatoria simple de hasta tres secciones.

Conglomerados

Cada IE es considerada un conglomerado, pues agrupa a un conjunto de secciones que, a su vez, agrupan a un conjunto de estudiantes. El muestreo de conglomerados surge como una necesidad práctica, económica y de eficiencia administrativa. Su mayor ventaja es que no se necesita contar con un marco muestral de cada elemento por analizar; solo es necesario tener el marco muestral a nivel de conglomerados (IE y secciones) para realizar la selección de los mismos (Lohr, 2021). Luego, en cada conglomerado seleccionado se puede obtener el listado de todas las secciones y seleccionar una muestra de estas.

Estratificación

En las evaluaciones muestrales del 2023 se realizó una estratificación porque, en primer lugar, esto produce un límite más pequeño para el error de estimación en relación con el que se generaría por una muestra aleatoria simple del mismo tamaño (Sheaffer et al., 2006). En segundo lugar, un muestreo estratificado nos protege de una mala muestra, en la que podríamos tener un desbalance en las cantidades muestreadas en ciertos estratos (Lohr, 2021).

Finalmente, en un contexto de tanta diversidad como el peruano, la variable de interés (rendimiento) toma diferentes valores en diferentes subpoblaciones (área, gestión o característica³), por lo que es importante reportar resultados para cada una de ellas. La combinación de estas tres subpoblaciones permite generar los siguientes estratos

³ Solo aplicable para primaria.

mutuamente excluyentes y colectivamente exhaustivos en los que cada elemento pertenece únicamente a un estrato.

En el caso de 2.º grado primaria, se conformaron los siguientes estratos:

- Estrato 1: Público - Urbano - Polidocente Completo.
- Estrato 2: Público - Urbano - Unidocente/Multigrado.
- Estrato 3: Público - Rural - Polidocente Completo.
- Estrato 4: Público - Rural - Unidocente/Multigrado.
- Estrato 5: Privado (Las IE privadas, casi en su totalidad, son urbanas y polidocentes).

La evaluación muestral para 4.º grado de primaria, además de los estratos antes mencionados, contó con 26 dominios⁴ que vendrían a ser cada Dirección Regional de Educación (DRE), considerando que las IE son administradas por estas instancias y que en ellas se forman los estratos anteriormente descritos. Por lo tanto, las DRE se combinaron con los estratos antes mencionados, teniendo casos como el de Lima Metropolitana y Callao, que solo consideraron los estratos 1 y 5, además del caso de Madre de Dios, donde no se tiene el estrato 2. Por ello, el total de estratos considerados fue 123.

En el caso de 2.º grado de secundaria, ya que todas las escuelas son polidocentes completas, los estratos son los siguientes:

- Estrato 1: Público – Urbano.
- Estrato 2: Público – Rural.
- Estrato 3: Privado.

Estos estratos fueron combinados con las 26 DRE, dando como resultados un total de 76 estratos, ya que Lima Metropolitana y Callao no tienen el estrato 2.

Tamaño de muestra

El diseño de muestra utilizado en la UMC es complejo y, como tal, a falta de una fórmula explícita se recurre al efecto diseño como lo sugiere Kish. Para el cálculo

⁴ La Evaluación Muestral en 2.º grado de primaria a diferencia de los otros grados no tiene desagregación por DRE.

del tamaño de muestra se utilizó como insumo el cálculo del efecto diseño (DEFF) de evaluaciones previas, y se estableció un margen de error de 7 puntos sobre la medida de cada área evaluada, así como un nivel de confianza del 95 %.

Con los insumos antes mencionados, se calculó el tamaño de muestra para cada DRE; dentro de cada una de ellas, se estratificó según lo descrito en la sección de estratificación. Como siguiente paso, se empleó la asignación de Neyman para determinar la cantidad de muestra en cada uno de los estratos, la cual es directamente proporcional a la variabilidad y el tamaño del mismo.

La asignación de Neyman, también llamada de mínima varianza, consiste en determinar los valores de muestra para cada estrato de forma que, para un tamaño de muestra fijo, la varianza de los estimadores sea mínima. La utilidad de esta afijación es mayor si hay grandes diferencias en la variabilidad de los estratos (Pérez, 2006).

2.4 Cobertura en la ENLA 2023

En las siguientes tablas se presenta el tamaño final de la muestra desagregado según gestión, área y característica para cada uno de los grados evaluados, así como la población evaluada y la cobertura alcanzada. En el 2023, la ENLA tenía programado evaluar 11 858 IE y aproximadamente 308 283 estudiantes entre todos los grados programados.

El porcentaje de estudiantes evaluados respecto de los programados a nivel nacional en las evaluaciones muestrales varía entre 85 % y 98 %, dependiendo del grado, mientras que en la Evaluación Censal de 4.º grado EIB las coberturas varían entre 68 % y 100 %. Dicha cobertura se calculó considerando la cantidad de estudiantes que llegan a la evaluación final del año escolar en cada IE, según la información del Siagie para el 2023.

En las siguientes tablas, se presenta el tamaño final de la muestra desagregada según gestión, área y característica para cada uno de los grados evaluados, así como la población evaluada y la cobertura alcanzada.

El porcentaje de estudiantes evaluados respecto de los estudiantes programados a nivel nacional varía entre el 93 % y el 95 % dependiendo del grado. Esta cobertura se calculó considerando la cantidad de estudiantes que llegan a la evaluación final del año escolar en cada IE, según la información del Siagie para el 2022.

Tabla 2.2 Cobertura de la ENLA 2P 2023 a nivel de IE y estudiantes.

2.º grado de Primaria	Programados		Evaluados		Cobertura %	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Nacional	438	11 383	435	10 570	99,3	92,9
Público-Urbano-Poli. Completo	125	5773	125	5 427	100	94,0
Público-Urbano-Unic/Multigrado	6	101	6	87	100	86,1
Público-Rural-Poli. Completo	52	929	52	873	100	94,0
Público-Rural-Unic/Multigrado	87	651	86	606	98,9	93,1
Privado	168	3929	166	3 577	98,8	91,0

Tabla 2.3 Cobertura de la ENLA 4P 2023 a nivel de IE y estudiantes.

4.º grado de primaria	Programados		Evaluados		Cobertura	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Nacional	5743	135 175	5742	128 095	100,0	94,8
Amazonas	221	3294	221	3162	100,0	96,0
Áncash	276	6647	276	6328	100,0	95,2
Apurímac	148	4091	148	4015	100,0	98,1
Arequipa	236	5595	236	5406	100,0	96,6
Ayacucho	220	4874	220	4748	100,0	97,4
Cajamarca	504	7167	504	6914	100,0	96,5
Callao	172	5821	172	5442	100,0	93,5
Cusco	213	5715	213	5557	100,0	97,2
Huancavelica	196	3069	196	2988	100,0	97,4
Huánuco	298	6054	298	5756	100,0	95,1
Ica	233	7390	233	7011	100,0	94,9
Junín	300	5856	300	5572	100,0	95,2
La Libertad	249	5626	249	5292	100,0	94,1
Lambayeque	211	5801	211	5397	100,0	93,0
Lima Metropolitana	153	4633	153	4299	100,0	92,8
Lima Provincias	214	5467	214	5163	100,0	94,4

4.º grado de primaria	Programados		Evaluados		Cobertura	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Loreto	314	7162	314	6377	100,0	89,0
Madre de Dios	77	2846	77	2661	100,0	93,5
Moquegua	66	2157	66	2098	100,0	97,3
Pasco	146	2679	146	2525	100,0	94,3
Piura	335	8145	335	7755	100,0	95,2
Puno	263	6011	263	5869	100,0	97,6
San Martín	333	7073	332	6762	99,7	95,6
Tacna	99	3014	99	2894	100,0	96,0
Tumbes	105	3876	105	3544	100,0	91,4
Ucayali	161	5112	161	4560	100,0	89,2

Tabla 2.4 Cobertura de la ENLA 2S 2023 a nivel de IE y estudiantes.

2.º grado de secundaria	Programados		Evaluados		Cobertura	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Total	3090	134 322	3088	125 852	99,9	93,7
Amazonas	162	5497	162	5082	100,0	92,5
Áncash	130	5722	130	5437	100,0	95,0
Apurímac	69	3401	69	3313	100,0	97,4
Arequipa	141	6269	141	6122	100,0	97,7
Ayacucho	118	4419	118	4258	100,0	96,4
Cajamarca	194	6963	194	6611	100,0	94,9
Callao	90	4726	90	4451	100,0	94,2
Cusco	142	6467	142	6299	100,0	97,4
Huancavelica	102	2779	102	2633	100,0	94,7
Huánuco	117	4663	117	4242	100,0	91,0
Ica	80	4357	80	4130	100,0	94,8
Junín	198	7597	198	7089	100,0	93,3
La Libertad	139	6407	139	5954	100,0	92,9
Lambayeque	111	5517	111	5237	100,0	94,9
Lima Metropolitana	195	9793	195	9289	100,0	94,9
Lima Provincias	107	4601	106	4376	99,1	95,1
Loreto	178	7646	178	6520	100,0	85,3

2.º grado de secundaria	Programados		Evaluados		Cobertura	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Madre de Dios	45	2735	45	2567	100,0	93,9
Moquegua	32	1659	32	1610	100,0	97,0
Pasco	92	3161	92	2808	100,0	88,8
Piura	158	8035	158	7606	100,0	94,7
Puno	130	5414	130	5289	100,0	97,7
San Martín	113	4902	112	4548	99,1	92,8
Tacna	73	3183	73	3103	100,0	97,5
Tumbes	49	2321	49	2075	100,0	89,4
Ucayali	125	6088	125	5203	100,0	85,5

Tabla 2.5 Cobertura de la ENLA 4P EIB 2023 a nivel de IE y estudiantes.

4.º grado de primaria EIB	Programados		Evaluados		Cobertura	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Total	2587	27 403	2572	24 120	99,4	88,0
Amazonas	221	2876	220	2342	99,5	81,4
Áncash	257	2463	257	2363	100,0	95,9
Apurímac	203	2101	202	2003	99,5	95,3
Arequipa	13	129	13	123	100,0	95,3
Ayacucho	213	2106	213	2033	100,0	96,5
Cajamarca	11	151	11	132	100,0	87,4
Cusco	340	3850	340	3613	100,0	93,8
Huancavelica	182	1801	182	1745	100,0	96,9
Huánuco	163	1613	163	1476	100,0	91,5
Junín	131	1410	131	1181	100,0	83,8
Lambayeque	49	525	49	485	100,0	92,4
Loreto	339	3830	330	2946	97,3	76,9
Madre de Dios	8	65	8	52	100,0	80,0
Moquegua	1	6	1	6	100,0	100,0
Pasco	31	262	30	182	96,8	69,5
Puno	220	1918	220	1852	100,0	96,6
San Martín	24	235	24	189	100,0	80,4
Ucayali	181	2062	178	1397	98,3	67,7

En relación con los niveles de inferencia de la ENLA 2023 en 2.º grado de primaria, se reportan resultados representativos en los siguientes niveles: a) nivel nacional; b) por sexo de los estudiantes; c) por gestión de las IE (estatales/no estatales); d) por ubicación geográfica de las IE (urbanas/rurales); e) por característica (polidocente completo/unidocente-multigrado); f) a nivel de DRE; y g) a nivel de gestión, área o característica dentro de la DRE. En 6.º grado de primaria los niveles de inferencia no incluyen a la DRE.

En 4.º grado de primaria, se pueden reportar resultados representativos en los siguientes niveles: a) nivel nacional; b) por sexo de los estudiantes; c) por gestión de las IE (estatales/no estatales); d) por ubicación geográfica de las IE (urbanas/rurales); e) por característica (polidocente completo/unidocente-multigrado); f) a nivel de DRE; y g) a nivel de gestión, área o característica dentro de la DRE. En el caso de 2.º grado de secundaria, los niveles de inferencia son similares a los de 4.º grado de primaria, sin considerar la característica de la IE, ya que en este nivel todas las escuelas son polidocentes completas.

En 4.º grado de primaria EIB, al tratarse de una evaluación censal, los resultados son representativos para cualquier nivel de desagregación, siempre y cuando se tome en cuenta el dato de su cobertura.

2.5 Pesos muestrales ENLA 2023

Los pesos son ponderadores que ayudan a corregir la distribución de la muestra en los estratos y a expandir la información muestral a la población. La idea principal de los pesos muestrales es que un individuo seleccionado con una probabilidad de inclusión de π representa a $1/\pi$ individuos en la población. Este valor de $1/\pi$ es llamado peso muestral (Lumley, 2010).

Peso por IE. El peso en la primera etapa es el inverso de la probabilidad de selección de un conglomerado (IE) en el interior de cada estrato. En el caso del muestreo proporcional al tamaño, se utilizan probabilidades conjuntas de selección de la IE. La inversa de esta probabilidad vendría a ser el peso de la IE; es decir:

$$pIE_{ij} = \frac{1}{p_{ij}}$$

donde:

p_{ij} = probabilidad de selección conjunta de la i -ésima IE en el j -ésimo estrato.

p_{IEij} = peso de la i -ésima IE en el j -ésimo estrato.

Peso por sección. Es el inverso de la probabilidad de selección de las secciones. Como ya se mencionó antes, la selección de secciones fue aleatoria simple. Si la IE seleccionada tiene tres o menos secciones, el peso será igual a 1.

$$psec_i = \frac{secIE_i}{seceva_i}$$

donde:

$psec_i$ = peso por sección en la i -ésima IE.

$secIE_i$ = total de secciones de la i -ésima IE.

$seceva_i$ = secciones evaluadas en la i -ésima IE.

Ajuste por estudiantes no evaluados en la sección. Este ajuste se realiza de manera separada para cada área evaluada.

$$a_{ki} = \frac{t_{ki}}{s_{ki}}$$

donde:

a_{ki} = Ajuste por estudiantes no evaluados en la k -ésima sección de la i -ésima IE.

t_{ki} = total de estudiantes matriculados en la k -ésima sección de la i -ésima IE.

s_{ki} = total de estudiantes evaluados en la k -ésima sección de la i -ésima IE.

Peso final. Se obtiene un peso para Lectura y otro para Matemática, los cuales están dados por la siguiente fórmula:

$$pf_{kij} = p_{IEij} \times psec_i \times a_{ki}$$

donde:

p_{fkij} = peso final en la k-ésima sección, en la i-ésima IE y el j-ésimo estrato.

p_{IEij} = peso de la i-ésima IE en el j-ésimo estrato.

$p_{secc i}$ = peso por sección en la i-ésima IE.

a_{ki} = ajuste por estudiantes no evaluados en la k-ésima sección de la i-ésima IE.

Ajuste por no respuesta en la ECE de 4.º grado de primaria EIB

En el caso de las evaluaciones censales no se calcula un peso muestral; sin embargo, es necesario el cálculo de un ajuste por no respuesta. Estos ajustes también son ponderadores que permiten expandir la información recolectada a toda la población, incluyendo aquellos que por algún motivo no pudieron participar.

Para el cálculo del ajuste por no respuesta se construyó una estratificación combinando el distrito, la característica de la escuela y el tipo de lengua (andina o amazónica). El cálculo del ajuste por no respuesta se realizó de la siguiente manera:

$$AJU_h = \frac{PROG_h}{EVA_h}$$

Donde:

AJU_h = Ajuste por no respuesta en el estrato h

EVA_h = Cantidad de estudiantes evaluados en el estrato h

$PROG_h$ = Cantidad de estudiantes matriculados en el estrato h

Operativo de campo

Capítulo 3

Capítulo 3

Operativo de campo

Considerando el Plan de Evaluaciones a implementarse en el 2023, se desarrollaron estrategias para alcanzar los objetivos de la ENLA 2023. Estas estrategias se plasmaron en las especificaciones técnicas de los Términos de Referencia de los servicios requeridos y en los manuales de procedimientos elaborados por la UMC.

Como parte del diseño operativo de la ENLA 2023, los procedimientos del operativo de campo se organizaron, implementaron y ejecutaron siguiendo tres principios fundamentales:

- 1. Confidencialidad:** Asegurar que solo el aplicador y los estudiantes tengan acceso a las pruebas.
- 2. Estandarización:** Asegurar que los instrumentos se apliquen en las mismas condiciones y se sigan los procedimientos de aplicación.
- 3. Probidad:** Promover la rectitud, honradez y honestidad en las acciones de todos los actores durante el proceso de evaluación.

Responsabilidades del INEI

Para la ejecución de la aplicación en campo de la ENLA 2023, se contó con el Instituto Nacional de Estadística e Informática (INEI) como operador logístico responsable a nivel nacional. Sus responsabilidades incluyeron:

- Resguardar los instrumentos y documentos desde su llegada a cada local de la sede de jurisdicción hasta su recojo.
- Contratar, capacitar y supervisar al personal de la Red Administrativa (RA) necesario para la operación de campo.
- Organizar logísticamente las capacitaciones de todos los niveles de la RA.
- Realizar el contacto previo y la actualización de datos de las instituciones educativas (IE).
- Aplicar los instrumentos en las IE y controlar su calidad en las sedes jurisdiccionales de acuerdo con los manuales de procedimientos emitidos por la UMC.
- Realizar un inventario computarizado de todos los instrumentos y documentos con código de barras en las sedes.

Responsabilidades del Minedu

El Minedu, por su parte, se responsabilizó de los siguientes procesos:

- Impresión, inventario y modulado de los instrumentos de aplicación (cuadernillos de pruebas, cuestionarios y documentos auxiliares).
- Impresión y modulado de los materiales de capacitación.
- Distribución de materiales de capacitación.
- Distribución y repliegue de materiales de aplicación.
- Consecución e implementación de equipos informáticos, muebles y servicios en las sedes jurisdiccionales a nivel nacional.

Además, el Minedu acompañó y monitoreó todas las etapas de la implementación de la evaluación para verificar el cumplimiento de los procesos y procedimientos. Por tal motivo, desplegó personal en las sedes jurisdiccionales, el cual realizó actividades de coordinación y monitoreo, acompañó a la RA en todos los procesos de ejecución de la Evaluación y apoyó en la solución de incidentes en las sedes jurisdiccionales. Asimismo, supervisó las actividades realizadas por los contratistas encargados de la impresión y distribución, para asegurar el cumplimiento de los Términos de Referencia y la ejecución en los plazos establecidos.

3.1 Organización territorial del operativo

Teniendo en cuenta el criterio de accesibilidad, la cantidad de instituciones educativas, grados y operativos programados, el país se organizó territorialmente en 28 sedes jurisdiccionales regionales, en las cuales se establecieron e implementaron 186 sedes jurisdiccionales.

En la tabla 3.1 se puede observar el número de IE programadas por operativo en cada jurisdicción regional de aplicación.

Tabla 3.1 Sedes jurisdiccionales regionales y distribución de IE según grado a evaluar

Sedes regionales	IE NO EIB		IE EIB		Total general
	2P Nacional	4P Regional	2S Regional	4P ECE SOLO L2	
AMAZONAS	15	336	213	229	793
ÁNCASH-CHIMBOTE	8	133	71	26	238
ÁNCASH-HUARAZ	7	144	57	231	439
APURÍMAC	2	148	69	212	431

Sedes regionales	IE NO EIB			IE EIB	Total general
	2P Nacional	4P Regional	2S Regional	4P ECE SOLO L2	
AREQUIPA	24	245	143	13	425
AYACUCHO	8	214	105	241	568
CAJAMARCA	26	392	144	5	567
CALLAO	11	173	91		275
CUSCO	12	186	133	335	666
HUANCAVELICA	5	184	96	185	470
HUANUCO	17	266	108	166	557
ICA	11	281	108	21	421
JUNÍN	23	305	200	132	660
LA LIBERTAD	31	247	139		417
LAMBAYEQUE	18	215	111	51	395
LIMA METROPOLITANA	307	212	213		732
LIMA PROVINCIAS	14	161	93	1	269
LORETO	10	224	122	97	453
MADRE DE DIOS	5	77	45	10	137
MOQUEGUA	2	66	32	1	101
PASCO	5	146	91	32	274
PIURA	29	343	161		533
PUNO	13	263	130	201	607
SAN MARTÍN-MOYOBAMBA	9	161	48	19	237
SAN MARTÍN-TARAPOTO	12	245	106	229	592
TACNA	2	99	73		174
TUMBES	4	99	47		150
UCAYALI	10	219	150	215	594
Total general	640	5784	3099	2652	12175

3.2 Conformación de la Red Administrativa de aplicación

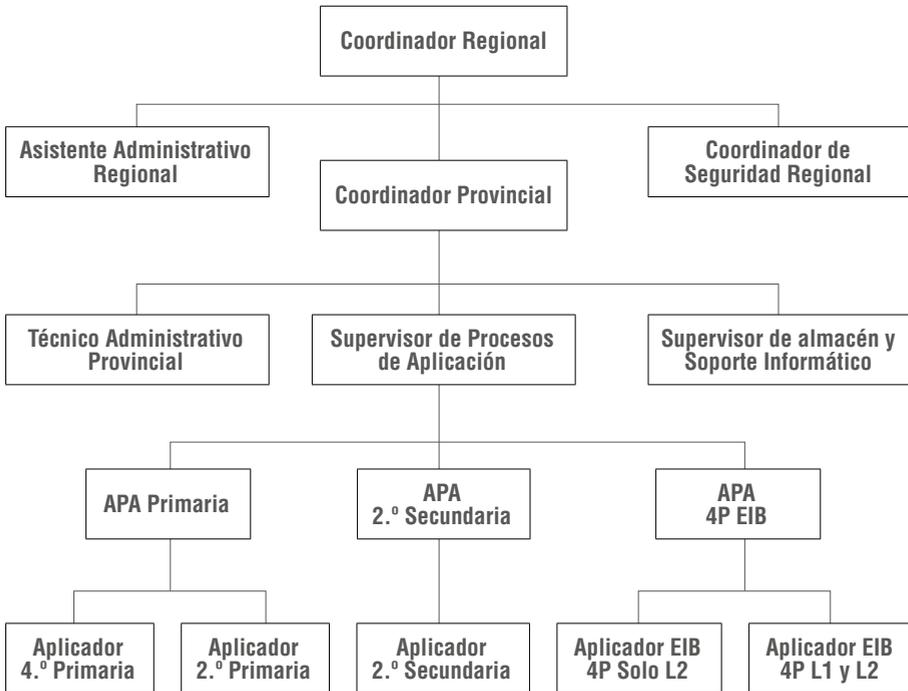
El personal necesario para ejecutar el operativo de campo estuvo organizado en una Red Administrativa (RA) conformada por los cargos mostrados en la tabla 3.2.

Tabla 3.2 Red Administrativa de aplicación

Nivel	Cargo en la RA de campo	Total RA 2023
Nivel IA	Supervisores nacionales	42
Nivel IB Nivel IB-1 Nivel IB-2	Coordinadores regionales (CR)	28
	Asistente administrativo regional	28
	Coordinador de seguridad regional	28
Nivel IC	Coordinadores provinciales (CP)	186
Nivel IC-1	Técnicos administrativos	186
Nivel IIA	Supervisor de procesos de aplicación (SPA)	49
Nivel IIB	Supervisores de almacén y soporte informático (SAS)	186
Nivel III Asistentes de procesos de aplicación	APA Primaria	773
	APA Secundaria	446
	APA EIB solo L2	206
	APA EIB con manejo de lenguas originarias	79
Nivel IV	Aplicadores 2P	1121
	Aplicadores 4P	10561
	Aplicadores 2S	7842
	Aplicadores 4P EIB SOLO L2 censal	5308
	Aplicadores de apoyo a estudiantes con discapacidad (aplicador ED)	1592

La Red Administrativa (RA) se organizó en campo tal como se muestra en la figura 3.1.

Figura 3.1 Organigrama de la Red Administrativa ENLA 2023



A continuación, se resumen las principales funciones de cada uno de los cargos:

- Los supervisores regionales velan por el cumplimiento de los procesos y procedimientos del operativo en la región asignada. Además, garantizan las condiciones materiales necesarias para la evaluación y aseguran la consecución de personal en los diferentes niveles de la RA.
- Los coordinadores regionales y provinciales son responsables de organizar, gestionar y dirigir el operativo, velar por el cumplimiento de los procedimientos y monitorear el desempeño de los miembros de la RA de su jurisdicción (región o provincia). Asimismo, son responsables de garantizar las condiciones materiales necesarias para el operativo y de asegurar la implementación de medidas de seguridad en el almacén de instrumentos.
- Los coordinadores de seguridad regional supervisan la efectividad de las medidas preventivas ante riesgos que puedan comprometer la integridad de la RA. Además, se encargan de supervisar la confidencialidad de los instrumentos de evaluación durante el operativo.

- Los supervisores de procesos de aplicación se encargan de la supervisión e implementación de convocar, seleccionar y capacitar a los asistentes de los procesos de aplicación. También, lideran y asisten técnicamente a los aplicadores.
- Los supervisores de almacén y soporte informático son responsables de gestionar la recepción, el almacenamiento, el inventario y el despacho del material de aplicación, así como del uso de los sistemas de información implementados para estas actividades.
- Los asistentes de procesos de aplicación son responsables de llevar a cabo la selección y capacitación de los aplicadores, velar por el cumplimiento de los procedimientos, supervisar la aplicación y realizar el contacto previo con las IE asignadas.
- Los aplicadores son responsables de la adecuada aplicación de los instrumentos en las IE.

Para cada uno de los niveles de la RA, el Minedu entregó manuales de aplicación y de funciones. En total, se elaboraron 18 documentos, entre manuales y protocolos, para el personal de la RA, atendiendo las particularidades de los operativos y grados evaluados.

- Manual del coordinador regional y provincial
- Manual del supervisor de procesos de aplicación
- Manual del supervisor de almacén y soporte informático
- Manual del usuario del sistema
- Manual del asistente de procesos de aplicación secundaria
- Manual del asistente de procesos de aplicación primaria
- Manual del asistente de procesos de aplicación EIB
- Manual del aplicador 2.º grado de primaria
- Manual del aplicador 4.º grado de primaria
- Manual del Aplicador 4P EIB L2
- Manual del aplicador 2.º grado de secundaria
- Protocolo del aplicador de apoyo a estudiantes con discapacidad visual — macrotipo

- Protocolo del aplicador de apoyo a estudiantes con discapacidad visual — prueba en braille
- Protocolo del aplicador de apoyo a estudiantes con discapacidad auditiva
- Protocolo del aplicador de apoyo a estudiantes con discapacidad física
- Protocolo del aplicador de apoyo a estudiantes con discapacidad del Trastorno del Espectro Autista
- Protocolo del aplicador de apoyo a estudiantes con discapacidad intelectual — lector
- Protocolo del aplicador de apoyo a estudiantes con discapacidad intelectual — lector inicial

3.3 Proceso de selección y capacitación del personal de la RA

El proceso de selección del personal estuvo a cargo del INEI y constó de tres etapas cancelatorias:

1. convocatoria y revisión de hojas de vida;
2. preselección; y
3. capacitación.

El Minedu estableció, en los Términos de Referencia del convenio con el INEI, el perfil profesional requerido para cada uno de los cargos. Por otro lado, basándose en la experiencia en campo y el análisis de las dificultades presentadas en años anteriores, se definieron los siguientes perfiles para los aplicadores:

Perfil del aplicador

OPCIÓN 1

Personal con mínimo una (1) experiencia demostrada en la aplicación de evaluaciones estandarizadas **cuyo desempeño no haya sido observado**. Debe cumplir:

- Ser como mínimo estudiante universitario, de instituto pedagógico o de carreras técnicas.

OPCIÓN 2

Personal sin experiencia en aplicación de evaluaciones estandarizadas.

Perfil académico:

- Egresado universitario o egresado técnico, excepto carreras vinculadas al cuidado personal u oficios; o
- estudiantes del último año de institutos pedagógicos; o
- estudiantes universitarios de los últimos dos años de carreras de humanidades.

Experiencia laboral:

- Docente con experiencia en aula, sin carga docente o administrativa en el sector estatal.

OPCIÓN 3

Personal sin experiencia en aplicación de evaluaciones estandarizadas

Perfil académico:

- Estudiantes universitarios o de institutos pedagógicos a partir del primer año; o
- estudiantes de carreras técnicas a partir del segundo año o egresados de CETPRO (excepto carreras vinculadas al cuidado personal u oficios).

Experiencia laboral:

- Con práctica sustentada en:
 - √ aplicación de instrumentos estandarizados de evaluación educativa; o
 - √ aplicación de instrumentos de recojo de información; o
 - √ participación en actividades con niños y adolescentes o en proyectos sociales.

Adicionalmente:

- Edad entre 19 y 55 años (salvo excepciones que deberán ser aprobadas por la UMC).
- Debe contar con habilidades para el buen manejo de grupos de niños y adolescentes.
- Buen nivel de lectura comprensiva.
- Adecuada dicción.
- Obligatoriedad de asistir a las sesiones de capacitación y repaso.

- Disposición a tiempo completo durante las fechas de capacitación y aplicación, así como para viajar o trasladarse y/o pernoctar en el lugar que se le designe.
- Obligatorio: que disponga de teléfono móvil.

Perfil adicional de aplicador

OPCIÓN 1

Con experiencia como Aplicador en operativos de la UMC (no necesita cumplir los otros requisitos del perfil):

- Experiencia demostrada como Aplicador, mínimo en uno (1) de los operativos de la UMC de 2015, 2016, 2018, 2019 o 2022, y cuyo su desempeño no haya sido observado.

OPCIÓN 2

Sin experiencia en la ECE-EM, pero con formación en:

- docencia con experiencia en aula, sin carga docente o administrativa en el sector estatal; o
- egresado universitario o egresado técnico, excepto carreras vinculadas al cuidado personal u oficios; o
- estudiantes de institutos pedagógicos; o
- estudiantes universitarios de carreras de humanidades; o
- estudiantes universitarios de otras carreras que no sean humanidades de los últimos tres años.

OPCIÓN 3

Sin experiencia en Operativos de la UMC, pero con prácticas afines en:

Perfil académico:

- Estudiantes universitarios o de institutos tecnológicos; o
- estudiantes de CETPRO (excepto carreras vinculadas al cuidado personal u oficios); o
- personal sin estudios técnicos o superiores.

Con práctica sustentada en:

- la aplicación de instrumentos estandarizados de evaluación educativa; o

- la aplicación de instrumentos de recojo de información; o
- participación en actividades con niños y adolescentes en aula o en proyectos sociales; o
- participación demostrada en proyectos de salud, educación, o alguna otra área social; o
- facilitación de actividades educativas o sociales.

Otras características a considerar para las OPCIONES 2 y 3:

- Edad entre 19 y 55 años (salvo excepciones que deberán ser aprobadas por la UMC).
- Habilidades para el buen manejo de grupos de niños y adolescentes.
- Buen nivel de lectura comprensiva.
- Adecuada dicción.
- Obligatoriedad de asistir a las sesiones de capacitación y repaso.
- Disposición a tiempo completo durante las fechas de capacitación y aplicación, así como para viajar al lugar que se le designe.
- Imprescindible: que disponga de teléfono móvil.

Perfil Aplicador EIB Solo L2

OPCIÓN 1

Con experiencia en Operativos de la UMC como aplicador (no necesita cumplir los otros requisitos del perfil):

- Experiencia demostrada como Aplicador en uno (1) de los operativos de la UMC de 2015, 2016, 2018, 2019 o 2022, y **cuyo su desempeño no haya sido observado.**

OPCIÓN 2

Sin experiencia en Operativos de la UMC, pero con formación en:

- docencia con experiencia en aula (sin carga docente o administrativa en el sector estatal); o
- egresado universitario o egresado técnico (excepto carreras vinculadas al cuidado personal u oficios); o
- estudiantes de institutos pedagógicos del último año; o

- estudiantes universitarios de carreras de humanidades de los últimos dos años.

OPCIÓN 3

Sin experiencia en Operativos de la UMC, pero con práctica sustentada en:

Perfil académico:

- Estudiantes universitarios o de Institutos Pedagógicos a partir del segundo año; o
- estudiantes de carreras técnicas a partir del primer año o egresados de CETPRO (excepto carreras vinculadas al cuidado personal u oficios).

Con práctica sustentada en:

- la aplicación de instrumentos estandarizados de evaluación educativa; o
- la aplicación de instrumentos de recojo de información; o
- participación en actividades con niños y adolescentes o en proyectos sociales.

Otras características a considerar para las OPCIONES 2 y 3:

- Edad entre 19 y 55 años.
- Deseable: Habilidades para el buen manejo de grupos de niños y adolescentes
- Buen nivel de lectura comprensiva
- Adecuada dicción
- Obligatoriedad de asistir a las sesiones de capacitación y repaso
- Disposición a tiempo completo durante las fechas de capacitación y aplicación, así como para viajar o trasladarse y/o pernoctar en el lugar que se le designe.
- Obligatorio: que disponga de teléfono móvil.

Aplicador de apoyo a estudiantes con discapacidad

Funciones: Apoyar en la aplicación de los instrumentos de evaluación a estudiantes con discapacidad.

Perfil mínimo (en orden de prioridad):

- Egresado de carrera de educación especial; o
- estudiante universitario de carreras de educación especial de los dos últimos años; o

- egresado de educación superior universitaria o titulado técnico con especialidad o diplomado en educación inclusiva; o
- egresados técnicos o estudiantes del último año de la carrera de educación, enfermería, fisioterapia y rehabilitación o carreras afines; o
- estudiantes universitarios de los dos últimos años o egresados de las carreras de psicología, educación, educación especial, trabajo social, sociología, antropología, comunicación social, medicina, enfermería, obstetricia, tecnología médica con mención en terapia física, terapia de lenguaje, terapia ocupacional o carreras afines.

Además, deberá contar con:

- Una (1) experiencia de trabajo con niños y adolescentes o población vulnerable (personas con discapacidad, poblaciones indígenas, adultos mayores o afines).
 - Para atención de estudiantes que presentan discapacidad auditiva (según registro RNEE), la persona deberá manejar la Lengua de Señas Peruana (LSP).
 - Para atención de estudiantes con TEA es deseable manejar el sistema de comunicación por intercambio de tarjetas.

Para la ENLA 2023, se programaron capacitaciones centralizadas, en Lima y capitales de región, y descentralizadas en cada sede jurisdiccional. Solo la capacitación de los aplicadores se realizó en sus respectivas sedes jurisdiccionales.

El objetivo de las capacitaciones fue desarrollar capacidades en los equipos de trabajo, familiarizándolos con los procedimientos e instrumentos de aplicación, y asegurar que todos reciban las mismas instrucciones en todo el país. Las capacitaciones de los primeros niveles de la RA se centraron en reforzar los procedimientos necesarios para llevar a cabo las actividades durante toda la implementación, en consonancia con las responsabilidades de cada nivel de la RA. En el caso de los aplicadores, se instruyó a los candidatos en la aplicación estandarizada de instrumentos, utilizando el Manual del Aplicador y los siguientes materiales elaborados específicamente para esta finalidad:

- fichas ópticas de práctica;
- papelógrafos;
- guiones de aplicación; y
- otros materiales pertinentes.

La capacitación para aplicadores tuvo una duración de dos días. Las aulas estuvieron conformadas por un máximo de 30 candidatos. Los aplicadores que atenderían a estudiantes con discapacidad recibieron medio día adicional de capacitación, según el tipo de discapacidad que iban a atender.

La metodología de la capacitación para todos los niveles incluyó:

- actividades de lectura individual y grupal;
- ejercicios prácticos para el uso de las fichas de recolección de información;
- actividades sobre manejo de procedimientos y documentos de aplicación; y
- simulaciones de las indicaciones para estudiantes en el aula de aplicación.

Se detallaron los indicadores de evaluación en cada una de las actividades propuestas. Al finalizar la capacitación, los candidatos rindieron una prueba de salida para evaluar su comprensión de los procedimientos centrales de la aplicación, así como su habilidad para manejar los documentos e instrumentos de aplicación. Además, se prepararon diseños de capacitación que fueron replicados en todas las aulas de capacitación a nivel nacional.

3.4 Características de los instrumentos y documentos utilizados

Se utilizaron los siguientes instrumentos y documentos de aplicación:

2.º grado de primaria

- Cuadernillos integrados de pruebas (un cuadernillo por día de evaluación)
- Ficha óptica de asistencia y respuestas (FOAR)
- Cuestionario para la familia
- Carta al director
- Cuestionario al director
- Cuestionario al docente
- Compromiso de confidencialidad para el docente
- Ficha del aplicador líder
- Registro de préstamo
- Ficha de verificación

4.º grado de primaria

- Cuadernillos integrados de pruebas (un cuadernillo por día de evaluación)
- Ficha óptica de asistencia y respuestas (FOAR)
- Cuestionario para la familia
- Carta al director
- Cuestionario al director
- Compromiso de confidencialidad para el docente
- Cuestionario al docente de matemática
- Cuestionario al docente de lectura
- Ficha del aplicador líder
- Registro de préstamo
- Ficha de verificación

4.º grado de primaria EIB solo L2

- Cuadernillos de pruebas
- Ficha óptica de asistencia y respuestas (FOAR)
- Cuestionario para la familia
- Carta al director
- Cuestionario al director
- Compromiso de confidencialidad para el docente
- Cuestionario al docente
- Ficha del aplicador líder
- Registro de préstamo
- Ficha de verificación

2.º grado de secundaria

- Cuadernillos integrados de pruebas (un cuadernillo por día de prueba)
- Cuestionario al estudiante día 2

- Cuestionario al estudiante día 3
- Fichas de respuestas para estudiantes
- Registro de estudiantes
- Carta al director
- Cuestionario al Director
- Carta al docente
- Cuestionario al docente de Comunicación
- Cuestionario al docente de Matemática
- Cuestionario al docente de Ciencias Sociales
- Cuestionario al docente tutor
- Compromiso de confidencialidad para el docente (por día de aplicación)
- Ficha del aplicador de sección
- Ficha del aplicador líder
- Registro de préstamo
- Ficha de verificación

Las cajas de aplicación que llegaron a cada sede jurisdiccional fueron inventariadas y almacenadas en el almacén del local de jurisdicción. Dicho local permaneció bajo llave y fue custodiado por la RA, con resguardo privado, hasta su retorno a Lima Metropolitana para su procesamiento.

3.5 Proceso de aplicación

La duración y las fechas de aplicación de cada operativo de la ENLA 2023 se muestran en la tabla 3.3:

Tabla 3.3 Duración y fechas de aplicación de la ENLA

Operativo	Duración de la aplicación	Fechas de aplicación	Criterio para asignar aplicadores
2P	2 días	21 y 22 de noviembre	Un aplicador líder por IE que requiera pernoctación de los aplicadores
			Un aplicador por sección programada
			Un aplicador de apoyo para secciones con 30 a más estudiantes
			Un aplicador para cada estudiante con discapacidad según registro de información UMC_ R-NEE
4P	2 días	21 y 22 de noviembre	Un aplicador líder por IE que requiera pernoctación de los aplicadores
			Un aplicador por sección programada
			Un aplicador de apoyo para secciones con 30 a más estudiantes
			Un aplicador para cada estudiante con discapacidad según registro de información UMC_ R-NEE
4EIB	1 día	21 de noviembre	Un aplicador líder por IE que requiera pernoctación de los aplicadores
			Un aplicador por sección programada
			Un aplicador de apoyo para secciones con 30 a más estudiantes
			Un aplicador para cada estudiante con discapacidad según registro de información UMC_ R-NEE
2S	3 días	21, 22 y 23 de noviembre	Un aplicador líder por IE que requiera pernoctación de los aplicadores
			Un aplicador por sección programada
			Un aplicador de apoyo para secciones con 30 a más estudiantes
			Un aplicador para cada estudiante con discapacidad según registro de información UMC_ R-NEE

A continuación, se presenta el resumen del proceso de aplicación:

Aplicación primaria (2.º y 4.º grados)

- I. Recepción de materiales. Los aplicadores de las IE que requieren pernoctación se desplazaron días antes del primer día de aplicación para llegar a tiempo a las fechas programadas. En las IE que no requieren pernoctación, el aplicador acudió cada día de aplicación a la sede para recoger los instrumentos de evaluación.

- II. Presentación ante el director de la IE y organización de la aplicación. El aplicador llegó temprano a la IE, se presentó ante el director, le entregó los documentos enviados por el Minedu y le solicitó la nómina de matrícula del grado a evaluar junto con la lista auxiliar de asistencia del docente para actualizar la lista de estudiantes a evaluar. El docente tutor firmó el compromiso de confidencialidad.
- III. Preparación de los materiales antes de la evaluación. El aplicador solicitó al director un espacio privado para preparar los cuadernillos y demás instrumentos. Solo y sin ayuda, el aplicador verificó los instrumentos que iba aplicar.
- IV. Preparación del aula e indicaciones. El aplicador organizó el mobiliario del aula, ubicó a los estudiantes según el orden de la Ficha óptica de asistencia y respuestas (FOAR), entregó los cuadernillos de las pruebas y dio lectura a las indicaciones de acuerdo con el guion de aplicación.
- V. Orden y duración de las sesiones de aplicación. El aplicador siguió el guion de aplicación y administró los cuadernillos siguiendo la secuencia y los tiempos indicados en la siguiente tabla:

Tabla 3.4 Orden y duración de las sesiones de aplicación de los grados de primaria

2.° grado de primaria		4.° grado de primaria	
Primer día	Segundo día	Primer día	Segundo día
Lectura (45 m)	Matemática (45 m)	Lectura (60 m)	Matemática (60 m)
Descanso (30 m)	Descanso (30 m)	Descanso (30 m)	Descanso (30 m)
Matemática (45 m)	Lectura (45 m)	Matemática (60 m)	Lectura (60 m)

- VI. Después de la aplicación, cada día, el aplicador se dirigió a un ambiente privado dentro de la IE para trasladar las respuestas de los cuadernillos a la FOAR. Luego, el director verificó que el traslado haya sido realizado correctamente, respetando los procedimientos. Al culminar la revisión, el aplicador contó la totalidad de cuadernillos y los lacró utilizando una bolsa de seguridad. El director presenció el conteo y firmó un acta de conformidad, dando fe de la cantidad de cuadernillos que retornaban al Minedu.
- VII. Al terminar las actividades previstas en la IE, el aplicador retornó inmediatamente a la sede para entregar todos los instrumentos de aplicación a su asistente de procesos de aplicación.
- VIII. Al finalizar el último día de la aplicación, el director o personal responsable de la IE dio conformidad a las condiciones de la aplicación en una declaración

jurada de aplicación, la cual fue firmada para cada sección evaluada. Para los casos en los que no se concretó la aplicación, se contó con una constancia de no aplicación, la cual detallaba el motivo de la no aplicación, el mismo que fue corroborado con una constancia de la UGEL.

Al culminar el trabajo en la IE, el aplicador se dirigió a la sede donde el asistente de procesos de aplicación verificó el material devuelto, teniendo en cuenta la cantidad y la calidad de la información recogida. Luego de este proceso, los asistentes de procesos de aplicación y los supervisores de almacén y soporte informático hicieron un inventario de los instrumentos. Las fichas ópticas fueron retiradas de sus cajas, clasificadas y ordenadas según correlativo e instrumento. Los paquetes de cuadernillos retornaron en sus mismas cajas.

Aplicación primaria EIB solo L2 (4.º grado)

- I. Recepción de materiales. Los aplicadores de las IE que requieren pernoctación se desplazaron días antes del primer día de aplicación para llegar a tiempo a las fechas programadas. En las IE que no requieren pernoctación, el aplicador acudió el día de aplicación a la sede para recoger los instrumentos de evaluación.
- II. Presentación ante el director de la IE y organización de la aplicación. El aplicador llegó temprano a la IE, se presentó ante el director, le entregó los documentos enviados por el Minedu y le solicitó la nómina de matrícula del grado a evaluar junto con la lista auxiliar de asistencia del docente para actualizar la lista de estudiantes a evaluar. El docente tutor firmó el compromiso de confidencialidad.
- III. Preparación de los materiales antes de la evaluación. El aplicador solicitó al director un espacio privado para preparar los cuadernillos y demás instrumentos. Solo y sin ayuda, el aplicador verificó los instrumentos que iba aplicar.
- IV. Preparación del aula e indicaciones. El aplicador organizó el mobiliario del aula, ubicó a los estudiantes según el orden de la Ficha Óptica de Asistencia y Respuestas (FOAR), entregó los cuadernillos de las pruebas y dio lectura a las indicaciones de acuerdo con el guion de aplicación.
- V. Orden y duración de las sesiones de aplicación. El aplicador siguió el guion de aplicación y administró los cuadernillos siguiendo la secuencia y los tiempos indicados en la siguiente tabla:

Tabla 3.5 Orden y duración de las sesiones de aplicación de 4.º grado de primaria EIB solo L2

Único día	Lectura en castellano como segunda lengua (40 m)
------------------	---

- VI. Después de la aplicación, el aplicador se dirigió a un ambiente privado dentro de la IE para trasladar las respuestas de los cuadernillos a la FOAR. Luego, el director verificó que el traslado haya sido realizado correctamente, respetando los procedimientos. Al culminar la revisión, el aplicador contó la totalidad de cuadernillos y los lacró utilizando una bolsa de seguridad. El director presenció el conteo y firmó un acta de conformidad, dando fe de la cantidad de cuadernillos que retornaban al Minedu.
- VII. Asimismo, el director o personal responsable de la IE dio conformidad a las condiciones de la aplicación en una declaración jurada de aplicación, la cual fue firmada para cada sección evaluada. Para los casos en los que no se concretó la aplicación, se contó con una constancia de no aplicación, la cual detallaba el motivo de la no aplicación, el mismo que fue corroborado con una constancia de la UGEL.

Al culminar el trabajo en la IE, el aplicador se dirigió a la sede donde el asistente de procesos de aplicación verificó el material devuelto, teniendo en cuenta la cantidad y la calidad de la información recogida. Luego de este proceso, los asistentes de procesos de aplicación y los supervisores de almacén y soporte informático hicieron un inventario de los instrumentos. Las fichas ópticas fueron retiradas de sus cajas, clasificadas y ordenadas según correlativo e instrumento. Los paquetes de cuadernillos retornaron en sus mismas cajas.

Aplicación secundaria (2.º grado de secundaria)

- I. Recepción de materiales. En las IE que no requerían pernoctación, el aplicador acudió cada día de aplicación a la sede para recoger los instrumentos de evaluación. Los aplicadores de las IE que requerían pernoctación se desplazaron con el tiempo necesario para iniciar la aplicación en las fechas programadas.
- II. Presentación ante el director de la IE y organización de la aplicación. El aplicador llegó temprano a la IE, se presentó ante el director, le entregó los documentos enviados por el Minedu y le solicitó la nómina de matrícula del grado a evaluar junto con la lista auxiliar de asistencia del docente. Se le solicitó que designara a un docente para el apoyo en la organización de los estudiantes, y se le solicitó que firmara el compromiso de confidencialidad.
- III. Preparación del aula e indicaciones. El aplicador organizó el mobiliario del aula, ubicó a los estudiantes según el orden del registro de estudiantes, entregó los cuadernillos de las pruebas y las fichas de respuestas, y luego dio lectura a las indicaciones de acuerdo con el guion de aplicación.
- IV. Orden y duración de las sesiones de aplicación. El aplicador administró los cuadernillos siguiendo la secuencia y los tiempos indicados en la siguiente tabla:

Tabla 3.6 Orden y duración de las sesiones de aplicación de 2.º grado de secundaria

Grado y operativo	Primer día	Segundo día	Tercer día
2.º grado de secundaria	Lectura (60 m)	Matemática (70 m)	Ciencias sociales (60 m)
	Descanso (30 m)	Descanso (30 m)	Cuestionario día 3 (30 m)
	Matemática (70 m)	Lectura (60 m)	
		Cuestionario día 2 (30 m)	

- V. Después de la aplicación, cada día, en un ambiente privado, el aplicador revisó las fichas de respuestas y organizó todos los instrumentos que debían ser devueltos al Minedu.
- VI. Al finalizar cada día de aplicación, el aplicador, en la dirección, contó la totalidad de cuadernillos y procedió a lacrarlos utilizando una bolsa de seguridad. El director, que presencié el conteo, firmó un acta de conformidad respecto de las condiciones de aplicación, dando fe de la cantidad de cuadernillos que retornaban al Minedu. Al terminar las actividades previstas en la IE, el aplicador retornó inmediatamente a la sede para entregar todos los instrumentos de aplicación a su Asistente de procesos de aplicación.
- VII. Al finalizar el último día de la aplicación, el director o personal responsable de la IE dio conformidad a las condiciones de la aplicación en una declaración jurada de aplicación, la cual fue firmada para cada sección evaluada. Para los casos en los que no se concretó la aplicación, se contó con una constancia de no aplicación, la cual detallaba el motivo de la no aplicación, el mismo que fue corroborado con una constancia de la UGEL.
- VIII. Al culminar el trabajo en la IE, el aplicador se dirigió a la sede donde el asistente de procesos de aplicación verificó el material devuelto, teniendo en cuenta la cantidad y la calidad de la información recogida. Luego de este proceso, los asistentes de procesos y los supervisores de almacén y soporte informático hicieron un inventario de los instrumentos. Las fichas ópticas fueron retiradas de sus cajas, clasificadas y ordenadas según correlativo e instrumento. Los paquetes de cuadernillos retornaron en sus mismas cajas.

Procesamiento de datos y análisis psicométrico

Capítulo 4

Para el procesamiento de datos y el análisis psicométrico, la UMC ejecuta un protocolo que contempla los lineamientos establecidos por la American Psychological Association (APA), la American Educational Research Association (AERA) y el National Council on Measurement in Education (NCME). De esta forma, garantiza la calidad de la información obtenida de sus evaluaciones.

4.1 Gestión y depuración de datos

La gestión y depuración de datos tiene como finalidad convertir en información digital consistente y confiable (base de datos) a aquella contenida en documentos físicos (fichas ópticas). Este proceso es realizado por una empresa especializada en digitalización y captura de datos, siguiendo especificaciones técnicas delineadas por la UMC. Al ser de suma importancia poder realizarlo en los plazos establecidos, la empresa responsable del proceso debe tener la capacidad logística necesaria para trabajar con una gran cantidad de datos.

La UMC se encarga de supervisar todos los procesos, subprocesos y actividades que realice la empresa de captura de datos, siendo esta la responsable de contratar el personal necesario para cumplir con los procesos y subprocesos necesarios. La UMC, por su parte, es responsable de la capacitación de este personal.

De manera general, este proceso consta de tres etapas: i) gestión física, ii) digitalización y captura de datos, y iii) depuración de bases de datos.

4.1.1 Gestión física

Esta etapa tiene como objetivo asegurar que la empresa contratada cuente con toda la documentación física a procesar, la cual debe estar completa y lista para la captura de datos. Por ello, todos los documentos y cajas utilizados en la evaluación tienen impreso un código de barras que los identifica. Además, todos los documentos vienen clasificados en cajas según su tipo. Por ejemplo: se cuenta con cajas distintas según se deba guardar fichas ópticas de respuestas, fichas ópticas de información complementaria de la aplicación, cuestionarios de factores asociados, cuadernillos o pruebas.

Se realizan las siguientes actividades:

- Recepción e inventario. Se reciben todas las cajas con los documentos físicos a procesar y se realiza el inventario en dos momentos: primero se revisa la cantidad

total de cajas y luego se revisa el contenido de cada caja. Este inventario se realiza escaneando el código de barras impreso en cada caja y en cada documento.

- Clasificación (de cuadernillos ED y de cuadernillos de la muestra de codificación). Se clasifican todos los documentos físicos según sus características.
- Preparación y loteo. Se quitan las grapas de los documentos y se limpian los residuos de borrador en las fichas ópticas, para que no se altere la imagen a digitalizar y, por ende, la captura de datos. Luego, se divide el total de documentos en paquetes pequeños (lotes) para gestionarlos más fácilmente.
- Almacenamiento. Se resguardan ordenadamente en un almacén todos los documentos físicos, para acceder a estos en caso sean requeridos en etapas siguientes.

Para asegurar la correcta gestión de los documentos físicos, la empresa contratada debe proporcionar un local con espacio suficiente durante todas las etapas del proceso, incluyendo un espacio exclusivo para el almacenamiento de todos los documentos garantizando su carácter confidencial.

4.1.2 Digitalización y captura de datos

Esta etapa tiene como objetivo digitalizar los documentos físicos y capturar los datos consignados en ellos. Implica las siguientes actividades:

- Digitalización. Consiste en escanear todos los documentos físicos inventariados. Los escáneres funcionan con un programa informático de control de calidad, el cual garantiza que la cantidad de documentos digitalizados es la misma que la de documentos inventariados.
- Captura de datos. Utilizando las imágenes digitalizadas y un programa informático desarrollado por la empresa contratada, se capturan los datos utilizando tres tecnologías: OMR (reconocimiento de marcas tipo burbujas), OCR (reconocimiento de caracteres impresos) e ICR (reconocimiento de caracteres escritos a mano).
- Control de calidad. Cuando el programa informático de reconocimiento de datos presenta un bajo nivel de confiabilidad, un grupo de personas se dedica a revisar y asegurar que los datos capturados correspondan a los datos consignados en la imagen del documento.
- Revisiones previas a la depuración. Finalizada la captura de datos, se debe asegurar que los archivos que se utilizarán en la depuración se hayan generado siguiendo las especificaciones dadas a la empresa contratada por la UMC. Las principales revisiones son las siguientes:
 - Cantidad de campos igual a la del diccionario de datos.

- Longitud y tipo de dato de todos los campos según diccionario de datos.
- Valores fuera de rango permitido según diccionario de datos.
- Campos de identificación completos.
- Correspondencia entre la data capturada y la imagen del documento de una muestra aleatoria de 50 casos de todo el archivo.

4.1.3 Depuración de datos

En esta etapa, se verifica que todos los campos de las bases de datos contengan respuestas que estén dentro de los márgenes permitidos, y que haya coherencia entre los campos y la información adicional que se maneja en la evaluación. Para ello, la UMC ha elaborado un Manual del depurador, el cual contiene el detalle de todas las revisiones a realizarse en cada campo de todas las bases de datos. Con este manual se capacita a todo el personal que participe en el proceso.

Durante esta etapa, algunas de las revisiones realizadas son las siguientes:

- Revisiones iniciales:
 - Correspondencia de registros según el reporte de inventario de instrumentos.
 - Registros únicos en los campos de identificación del instrumento (como el código de barras).
 - Correspondencia entre el código modular y el correlativo Minedu (según padrón de IE).
 - Hojas faltantes en cada ficha óptica.
- Revisiones principales (por tipo de documento):
 - Caracteres no válidos en los DNI de los estudiantes.
 - Duplicados en los DNI de los estudiantes.
 - Caracteres no válidos en los nombres y apellidos de los estudiantes.
 - Duplicados en los apellidos y nombres de los estudiantes.
 - Caracteres no válidos en los correlativos de los estudiantes.
 - Duplicados en los correlativos de los estudiantes.
 - Caracteres no válidos en el sexo de los estudiantes.
 - Caracteres no válidos en la lengua materna de los estudiantes.

- Caracteres no válidos en la discapacidad de los estudiantes.
 - Caracteres no válidos en los números de forma de los cuadernillos.
 - Caracteres no válidos en las respuestas de los estudiantes.
 - Inconsistencias en las respuestas y la asistencia de los estudiantes.
- Revisiones finales:
 - Repetir todas las revisiones iniciales.
 - Contrastar con el listado de IE aplicadas, el cual es proporcionado por el equipo de la UMC encargado del trabajo de campo.
 - Identificar valores fuera del rango permitido.

4.2 Estrategias de análisis psicométrico

4.2.1 Modelo Rasch para ítems dicotómicos

Para la Evaluación Nacional de Logros de Aprendizaje (ENLA) 2023, con excepción de la población en IE EIB, se utilizó el modelo Rasch para ítems dicotómicos (Bond et al., 2021; Wright y Stone, 1998), el cual se centra en el análisis de cada ítem que contempla dos posibles respuestas (correcto/ incorrecto). Específicamente, se centra en la interacción entre un ítem y una persona. Este modelo establece la probabilidad de respuesta de una persona ante un ítem en términos de la diferencia entre la medida de rasgo o habilidad latente de la persona (θ) y la medida de dificultad del ítem utilizado (β)¹. Por este motivo, se le denomina usualmente modelo de un parámetro (Hambleton et al., 1991).

En los modelos Rasch, la habilidad de las personas y las dificultades de los ítems se ubican en una misma métrica cuya interacción permite predecir el desempeño de una persona frente a un ítem cualquiera, así como identificar respuestas inesperadas (Smith y Kramer, 1989). El objetivo central del análisis Rasch es construir una escala conformada por los ítems ordenados según su dificultad. Ello implica que, a mayor habilidad, la persona presentará mayor probabilidad de acertar y, por lo tanto, obtendrá un mayor número de respuestas correctas. Es importante tener en cuenta que la medida estimada de la persona no es igual que el puntaje directo que posee (número de ítems correctos); este puntaje es solo un insumo a partir del cual se construye la medida Rasch. La relación entre la habilidad y la dificultad puede graficarse por medio de las curvas características del ítem (CCI), las cuales brindan información concreta sobre la probabilidad de respuesta de una persona ante dicho

¹ En algunos textos, se utiliza la letra B para referirse a la habilidad de las personas y la letra D para referirse a la dificultad de los ítems.

ítem. Al trazar estas curvas, en caso de tener ítems dicotómicos, se establecen las siguientes relaciones:

$$\theta > \beta; p(X = 1|\theta, \beta) \in]0,5; 1,0[$$

$$\theta < \beta; p(X = 1|\theta, \beta) \in [0,0; 0,5[$$

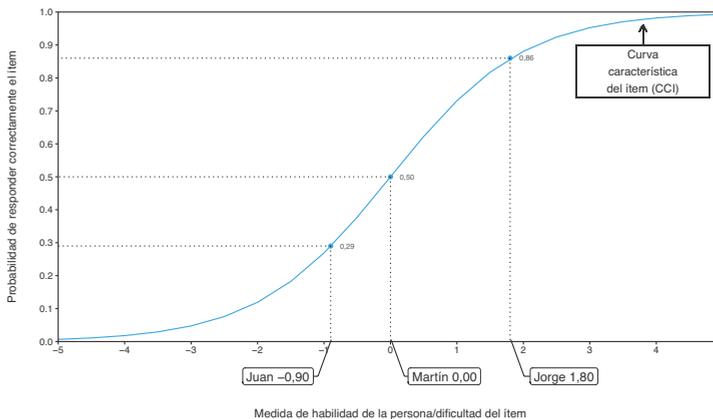
$$\theta = \beta; p(X = 1|\theta, \beta) = 0,5$$

En el primer caso, si la habilidad de la persona es mayor que la dificultad del ítem, la probabilidad de responder correctamente el ítem es mayor que 0,5 (50 %). En el segundo caso, si la habilidad de la persona es menor que la dificultad del ítem, la probabilidad de responder correctamente dicho ítem es menor que 0,5 (50 %). Finalmente, en el tercer caso, si la habilidad de la persona es igual a la dificultad del ítem, la probabilidad de responder correctamente dicho ítem es igual que 0,5 (50 %). Como señala Ingebo (1997), de esta manera se puede comprobar empíricamente la teoría de que los estudiantes con mayores conocimientos tienen mayores probabilidades de responder correctamente una pregunta, a diferencia de los estudiantes con menores conocimientos. Matemáticamente, la CCI se grafica con la siguiente función:

$$P(X_{ni} = 1|\theta_n, \beta_i) = \frac{e^{\theta_n - \beta_i}}{1 + e^{\theta_n - \beta_i}}$$

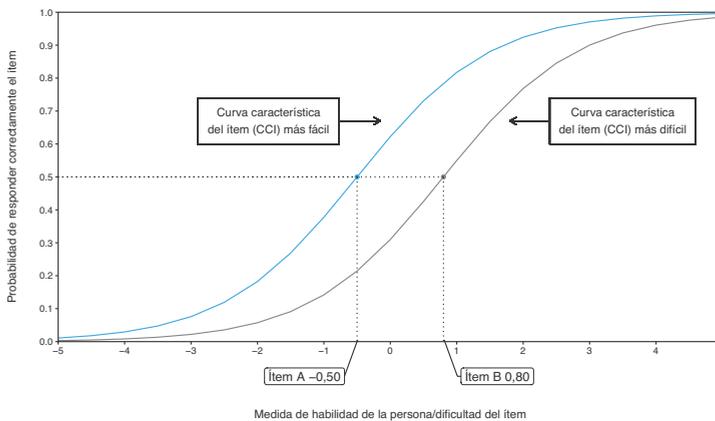
Donde θ representa el rasgo latente de cada persona n , y β representa la medida de dificultad de cada ítem i . Esta relación entre la habilidad de una persona y la dificultad de un ítem se muestra en la figura 4.1.

Figura 4.1 Curva característica del ítem y probabilidades de tres personas de responderlo correctamente



Como se observa, Juan tiene una habilidad de $-0,90$, por lo que su probabilidad de acertar este ítem es de $0,29$; es decir, lo más probable es que falle y obtenga 0 puntos. En cambio, Jorge, cuya habilidad se ha estimado en $1,80$, tiene una probabilidad de $0,86$ de acertar este ítem; por lo tanto, es más probable que lo responda correctamente y reciba 1 punto. Finalmente, Martín tiene una habilidad igual a la dificultad del ítem; por eso, se afirma que tiene la misma probabilidad de acertar o de fallar el ítem. Al comparar dos o más curvas características de ítems se puede decir que, mientras más hacia la derecha se desplaza una curva, más difícil es el ítem. Además, la dificultad de un ítem se expresa numéricamente como el valor en la escala de habilidad que posee una probabilidad del 50 % de acertar dicho ítem. En la figura 4.2, el ítem más fácil (A) tiene una dificultad de $-0,50$ y el más difícil (B), de $0,80$.

Figura 4.2 Curva característica de dos ítems y comparación de la dificultad de ambos



Como ya se señaló, θ representa la medida verdadera del rasgo latente de una persona n . Es decir, sería su medida de habilidad si se la pudiese evaluar en condiciones óptimas y con todos los ítems que potencialmente pueden usarse para medir dicho rasgo latente. No obstante, como esto no es posible en términos empíricos, lo que se tiene es una estimación de dicha habilidad, representada por $\hat{\theta}$. Por su parte, β representa el parámetro de dificultad de un ítem i a nivel poblacional, por lo cual, a nivel de muestra, se lo representa como b . De esta manera, el modelo Rasch puede ser expresado del siguiente modo:

$$P_{ni} = P(X_{ni} = 1 | \hat{\theta}_n, b_i) = \frac{e^{\hat{\theta}_n - b_i}}{1 + e^{\hat{\theta}_n - b_i}}$$

La dificultad del ítem (b) es un estimado, pues está sujeta a un grado de incertidumbre que se expresa mediante el error estándar de la dificultad del ítem (Wilson, 2005). Este error estándar puede ser calculado mediante la fórmula:

$$e. e. (b_i) = \frac{1}{\sqrt{\sum_{n=1}^{I-1} N_n P_{ni} (1 - P_{ni})}}$$

4.2.2 Modelo Rasch de crédito parcial

Mientras el modelo Rasch para ítems dicotómicos se aplicó a dicho tipo de ítems, para los politómicos de respuesta categóricas ordenadas se recurrió al modelo de créditos parciales (Masters, 1982). Este se deriva de las pruebas de opción múltiple: algunas respuestas incorrectas reciben crédito o puntaje parcial porque indican algún conocimiento sobre la respuesta apropiada (Wright, 1999). Además, este modelo especifica que cada ítem tiene su propia estructura de calificación. Según Fox (1999), es una generalización del modelo Rasch para ítems dicotómicos aplicable en pruebas cuyos ítems pueden variar en el número de alternativas correctas y en la cantidad de opciones de respuesta. Este modelo supone que el proceso de resolución de un ítem involucra la aplicación secuencial de un conjunto de pasos. Los créditos parciales dados equivalen al número de pasos que deben completarse antes de obtener la respuesta final del problema (Masters, 1982). Por ejemplo, supongamos que un estudiante debe hallar el valor de x en el siguiente ítem matemático:

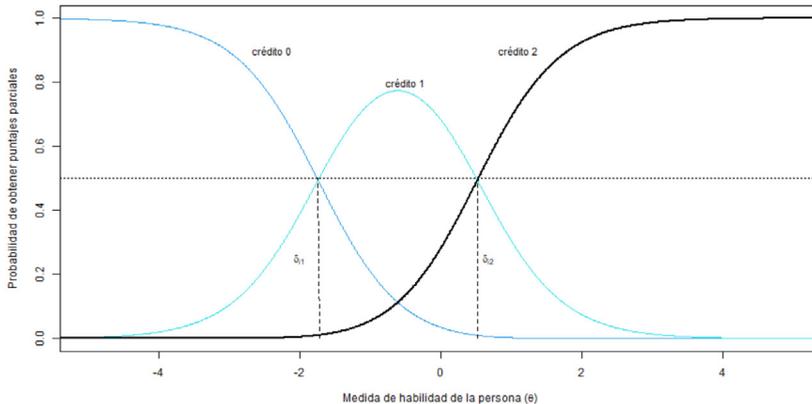
$$\sqrt{\frac{196}{4}} = x$$

Existe una secuencia de dos pasos para la solución de este ítem: (1) se divide 196 entre 4, lo que da como resultado 49; y (2) se saca la raíz cuadrada de 49 para obtener el resultado final, que es 7. Si el estudiante no puede resolver siquiera la división, recibe 0 puntos. Si es capaz de resolver la división, pero no la raíz cuadrada, recibe 1 punto. Si, además de resolver la división, obtiene la raíz cuadrada, recibe 2 puntos.

El modelo asume que puntajes o créditos superiores son más probables de ser alcanzados en ítems con baja dificultad que en ítems con alta dificultad. Además, los estudiantes con mayores niveles del rasgo latente tienen mayor probabilidad de recibir el crédito total para el ítem que los estudiantes con menores niveles del rasgo latente. Ambas propiedades deben mantenerse para todas las personas, ítems y categorías de respuesta. Masters (1982) indica que, cuando estas propiedades se cumplen, los puntajes directos obtenidos pueden ser transformados para lograr una escala de intervalo.

Una manera de graficar el modelo de crédito parcial es mediante las curvas de categorías de respuesta (Embretson y Reise, 2000). Estas curvas muestran la relación entre la cantidad del rasgo latente evaluado y la probabilidad de obtener puntajes parciales. Además, indican en qué lugar del continuo del rasgo latente, las respuestas a una categoría son más probables que las respuestas a otra categoría. La figura 4.3 presenta un ejemplo de este tipo de curvas.

Figura 4.3 Curva de categorías de respuestas con umbrales Rasch-Andrich



Los parámetros δ_{ij} se encuentran en la intersección de las curvas de las categorías de respuesta. Se les denomina umbrales Rasch-Andrich (Linacre, 2001) y representan el valor en la escala del rasgo latente en el cual es igualmente probable recibir un puntaje o el inmediatamente adyacente. En algunos casos (Linacre, 2001), δ_{ij} es descompuesto en dos términos, D_i y F_{ij} , de tal manera que

$$\delta_{ij} = D_i + F_{i1}$$

y

$$\delta_{ij} = D_i + F_{i2}$$

En este ejemplo, el primer umbral con un valor de $-1,74$ representa la medida en la cual es igualmente probable recibir 0 y 1 punto en el ítem. A partir de la figura 4.3, además, se pueden señalar tres situaciones para una persona: si esta tiene una medida del rasgo latente menor a $-1,74$, lo más probable es que reciba 0 puntos (crédito 0) en este ítem; si su medida se encuentra entre $-1,74$ y $0,52$, lo más probable es que reciba 1 punto (crédito 1); y, si su medida de habilidad es mayor que $0,52$, lo más probable es que reciba 2 puntos (crédito 2).

Para expresar matemáticamente el modelo de crédito parcial, se debe partir del supuesto de que los ítems pueden recibir puntuaciones desde 0 hasta un máximo (2, 3, etc.). Ello se expresa del siguiente modo:

$$X_{ni} = x \in (0,1, \dots, m_i)$$

Donde:

n : identificador (número) de persona

i : identificador (número) de ítem

X_{ni} : puntaje obtenido por una persona n en un ítem i

x : puntaje observado

m_i : crédito máximo posible en un ítem i

La probabilidad de obtener cierto puntaje en un ítem viene dada por el modelo de crédito parcial, el cual se representa usando una adaptación de la notación original propuesta por Masters (1982):

$$P(X_{ni} = x) = \frac{\exp \sum_{j=0}^x (\theta_n - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^k (\theta_n - \delta_{ij})}$$

Donde:

j : número del umbral (0 a m_i-1)

δ_{ij} : valor del umbral j , del ítem i

Es importante considerar que, por conveniencia computacional, el valor δ_{i0} se elige de tal manera que

$$\sum_{j=0}^k (\theta_n - \delta_{ij}) = 0$$

Supongamos que un ítem i tiene tres puntajes posibles, es decir, $X_{ni}=x \in (0,1,2)$. A partir de ello, se puede observar que $m_i=2$. Por lo tanto, las probabilidades de obtener 0, 1 o 2 puntos se pueden expresar de la siguiente manera:

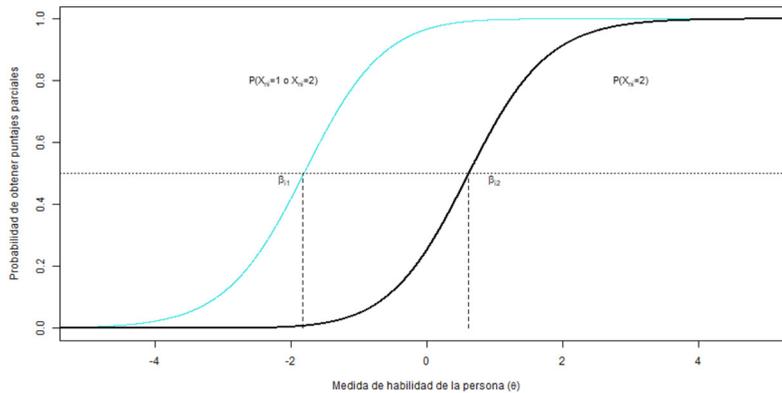
$$P(X_{ni} = 0) = \frac{\exp(0)}{\exp(0) + \exp(0 + (\theta_n - \delta_{i1})) + \exp(0 + (\theta_n - \delta_{i1}) + (\theta_n - \delta_{i2}))}$$

$$P(X_{ni} = 1) = \frac{\exp(0 + (\theta_n - \delta_{ij}))}{\exp(0) + \exp(0 + (\theta_n - \delta_{i1})) + \exp(0 + (\theta_n - \delta_{i1}) + (\theta_n - \delta_{i2}))}$$

$$P(X_{ni} = 2) = \frac{\exp(0 + (\theta_n - \delta_{ij}) + (\theta_n - \delta_{ij}))}{\exp(0) + \exp(0 + (\theta_n - \delta_{i1})) + \exp(0 + (\theta_n - \delta_{i1}) + (\theta_n - \delta_{i2}))}$$

Como se señaló, el modelo de crédito parcial intenta captar diferentes niveles de conocimiento frente a un mismo ítem. Por ejemplo, en un ítem de dos pasos, se modela la probabilidad de obtener más de 0 puntos y, luego, la probabilidad de obtener más de 1 punto. Este proceso se puede representar con dos curvas características del ítem, una para cada paso, como se muestra en la figura 4.4.

Figura 4.4 Curva de categorías de respuestas con umbrales Rasch-Thurstone



La curva de la izquierda se refiere a la probabilidad de recibir 1 o 2 puntos en lugar de 0. La curva de la derecha indica la probabilidad de recibir 2 puntos frente a 1 o 0 puntos. Es decir, una persona con una medida de rasgo latente de $-1,83$ tiene una probabilidad igual al 50 % de superar el primer paso (tener más de 0), mientras que una persona con una medida de $0,61$ tiene una probabilidad del 50 % de superar el segundo paso (más de 1 punto). A dichos valores ($-1,83$ y $0,61$), se les denomina umbrales Rasch-Thurstone (Linacre, 2001).

4.2.3 Análisis de ítems

El buen ajuste a un modelo es una parte importante en cualquier análisis estadístico; en el caso del análisis Rasch, es una parte esencial (Wilson, 2005). Las relaciones datos-modelo son muy diferentes de lo que usualmente se aplica en el análisis estadístico tradicional, en el cual el desajuste de los datos respecto del modelo suele dar lugar a la desestimación del modelo estadístico. En el caso del análisis Rasch, los valores que no ajustan (*misfit*) conducen al rechazo de los datos, ya que no cumplen con los requisitos del modelo. Por lo tanto, se debe entender el modelo Rasch como un modelo prescriptivo en el que se investiga cómo los datos se ajustan al modelo y no cómo el modelo se ajusta a los datos (Bond et al., 2021). Como señala Wilson (2005), el ajuste de los ítems al modelo Rasch se fundamenta en dos datos estadísticos: el outfit y el infit, cuyo cálculo se basa en los residuos cuadráticos estandarizados del modelo (Z_{ni}^2). A partir de las propuestas de diversos autores

(Bond et al., 2021; Wilson, 2005; Wright y Linacre, 1994), se exponen brevemente el significado y el cálculo de ambos indicadores de ajuste.

El outfit (*outlier sensitive fit statistic*) es más sensible al comportamiento inesperado alejado de la medida de una persona. Se basa en la suma de cuadrados de los residuales estandarizados (Z_{ni}^2), de tal manera que

$$Outfit_i = \frac{1}{N} \sum_{n=1}^N Z_{ni}^2$$

El infit (*information weighted fit statistic*) está afectado básicamente por respuestas no esperadas a ítems calibrados cerca de la medida de la persona. Su cálculo se basa en la siguiente fórmula:

$$Infit_i = \frac{VAR(P_{ni})}{\sum_{n=1}^N VAR(P_{ni})}$$

Como $\sum_{n=1}^N R_{ni}^2 = VAR(P_{ni}) Z_{ni}^2$, se puede redefinir el infit como una media cuadrática ponderada de residuales (Z_{ni}^2) que es sensible a patrones de respuesta irregulares.

$$Infit_i = \sum_{n=1}^N \omega_{ni} Z_{ni}^2$$

En la fórmula anterior, los residuales están ponderados por sus varianzas individuales.

$$Infit_i = \frac{\sum_{n=1}^N R_{ni}^2}{\sum_{n=1}^N VAR(P_{ni})}$$

En cuanto a la interpretación de los valores infit y outfit, puede señalarse como ejemplo que un valor de ajuste de 1,25 indica que los datos presentan un 25 % más de "ruido" que el modelado.

Además de los índices antes mencionados, se puede calcular la correlación de Pearson entre la respuesta al ítem y la medida estimada para cada persona (ptme). A continuación, se presenta la fórmula para calcularlo (Linacre, 2024):

$$ptme = \frac{\sum_{n=1}^N [(X_n - \sum_{m=1}^N X_m/N)(\hat{\theta}_n - \sum_{m=1}^N \hat{\theta}_m/N)]}{\sqrt{\sum_{n=1}^N [(X_n - \sum_{m=1}^N X_m/N)]^2 \sum_{m=1}^N [(\hat{\theta}_n - \sum_{m=1}^N \hat{\theta}_m/N)]^2}}$$

Donde X_1, \dots, X_N son las respuestas de las personas a los ítems, y $\hat{\theta}_1, \dots, \hat{\theta}_N$ son las medidas estimadas para cada persona.

Siguiendo las recomendaciones de Linacre (2024), se espera que los ítems con un ajuste adecuado al modelo Rasch tengan valores infit y outfit entre 0,50 y 1,50; sin embargo, dentro de este grupo, se prefieren los valores entre 0,70 y 1,30 (Wright y Linacre, 1994). Algunos autores (Schulz, 1990) señalan que debe prestársele mayor atención al infit, ya que es menos sensible a las variaciones en el tamaño de la muestra utilizada para calibrar los ítems. Además, se esperan correlaciones ítem-medida (*ptme*) positivas (Linacre, 2024).

Asimismo, se calcula el promedio de puntaje por ítem (p) que corresponde a la tasa de acierto en un ítem dicotómico y al promedio ponderado de las respuestas en el caso de un ítem de crédito parcial.

4.2.4 Evidencias de validez vinculadas a la estructura interna

La evidencia de validez basada en la estructura interna requiere, en primer lugar, que se tenga claridad conceptual sobre dicha estructura dado el constructo de interés (Wilson, 2005). En este sentido, un tipo de evidencia está relacionado con la unidimensionalidad de las medidas derivadas de la aplicación de la prueba. El concepto de unidimensionalidad implica que un solo rasgo latente (o constructo) se encuentra en la base de un conjunto de ítems (Hattie, 1985). En otras palabras, un instrumento es unidimensional si las respuestas son producidas sobre la base de un único atributo. Wright y Linacre (1989) señalan que, en la práctica, ningún instrumento puede ser perfectamente unidimensional. Por ello, lo que se busca es tener instrumentos que, en esencia, muestren unidimensionalidad. Por ejemplo, muchos factores como la motivación, la ansiedad y la velocidad de respuesta tienen un impacto sobre el desempeño de una persona en un conjunto de ítems (Hambleton et al., 1991). No obstante, lo importante es que un instrumento de medida represente con sus puntuaciones un solo factor dominante. Con esto, lo que se quiere lograr es que la mayor cantidad de la varianza observada en las respuestas a los ítems sea explicada por un solo atributo latente (Embretson y Reise, 2000). Esto se suele analizar aplicando un análisis factorial, el cual debe ser diferenciado del análisis de componentes principales (Preacher y MacCallum, 2003). El primero trata de describir la estructura latente de un conjunto de variables usando factores que expliquen la varianza común; el segundo busca reducir las variables a un conjunto menor de componentes enfocándose en la varianza total (Park et al., 2002; Preacher y MacCallum, 2003).

En el caso de los modelos Rasch, lo usual es efectuar un análisis de componentes principales de los residuos estandarizados luego de ajustar el modelo Rasch unidimensional (Linacre, 2024). Este análisis de los residuos sirve para detectar cualquier varianza que quede en los datos luego de extraer la dimensión principal

(Bond et al., 2021), lo cual se realiza tratando de modelar la varianza que no es explicada por el constructo que se pretende medir. La idea básica detrás de este análisis es que, si las respuestas cumplen adecuadamente con el requerimiento de unidimensionalidad, no debería haber patrones significativos en los residuos. En otras palabras: no debería haber una variable latente adicional, distinta de la variable latente principal, que haya tenido un impacto en las respuestas a los ítems (Wind y Hua, 2022).

Linacre (1998) ha demostrado que este tipo de residuos sirve para detectar mejor la presencia de dimensiones secundarias que otros tipos de residuos, como los no estandarizados o los logarítmicos. Además, este mismo autor señala que los resultados obtenidos con el análisis de componentes principales son similares a los del análisis factorial, pero prefiere el análisis de componentes principales por su rigurosa base matemática.

Asimismo, Linacre (2024) afirma que, si bien no existen parámetros absolutos para interpretar los resultados del análisis de componentes principales de los residuos, si el primer autovalor contiene menos del 5 % de varianza, o si su valor es menor que 3, entonces no existen serias evidencias que atenten contra el supuesto de unidimensionalidad y, por lo tanto, los datos se pueden analizar adecuadamente utilizando modelos Rasch. Asimismo, se está frente a datos unidimensionales cuando la cantidad de varianza empírica es similar a la cantidad de varianza predicha por el modelo.

4.2.5 Confiabilidad y consistencia de la clasificación

El índice de confiabilidad de la separación de las personas (R_p) es el coeficiente utilizado para las pruebas aplicadas por la UMC. Este índice es análogo al de consistencia interna α (Cronbach, 1951), pero produce mejores estimaciones, pues los valores numéricos son lineales si los datos se ajustan al modelo Rasch aplicado (Wind y Hua, 2022). Además, utiliza la varianza de error promedio de la muestra en lugar de la varianza de error de una persona promedio (Schumacker y Smith, 2007). El coeficiente sirve para indicar la capacidad de las medidas de un test para diferenciar las cantidades del rasgo latente que poseen los evaluados (Wright y Masters, 1982). En ese sentido, indica la replicabilidad del ordenamiento de las personas según su medida de habilidad si se les da otro conjunto de ítems que miden el mismo constructo (Bond et al., 2021).

Un índice menor a 0,50 indica que las diferencias entre las medidas son producidas principalmente por el error de medición (Fisher, 1992). Sobre los valores mínimos aceptables de los coeficientes de confiabilidad, Charter (2003) ha realizado una revisión de numerosas investigaciones que proponen diferentes niveles mínimos. En ese estudio, encontró mucha variabilidad; asimismo, observó valores propuestos con los diversos métodos para obtener la confiabilidad, los cuales oscilan entre 0,60 y

0,95. A pesar de esta gran variabilidad, un estándar mínimo aceptable que aparece con frecuencia en la literatura es el de 0,70, señalado por Nunnally y Bernstein (1995).

Estos coeficientes también pueden ser expresados como índices de separación de personas. Estos índices se refieren a la dispersión de los datos medidos como el número de errores estándar que separan a las personas (Schumacker y Smith, 2007).

Otros indicadores importantes para el análisis son los de consistencia y precisión de la clasificación. Por un lado, la *consistencia* de la clasificación es el grado de acuerdo en dos aplicaciones independientes o paralelas de un instrumento de medición. Se espera que el alumno sea clasificado en la misma categoría al repetirse la evaluación. Por su parte, la precisión de la clasificación implica el grado en el cual la clasificación observada coincide con la clasificación verdadera (Kim et al., 2006). En las evaluaciones realizadas, se utilizó el método de Rudner (una aproximación de tipo individual) pues calcula la consistencia de clasificación para cada persona y luego las promedia (Lee, 2010). Además, asume que los errores de estimación se distribuyen normalmente. A medida que aumenta el número de ítems, dadas las propiedades del estimador $\hat{\theta}$ basado en un método de máxima verosimilitud, este supuesto es más plausible.

4.3 Resultados del análisis psicométrico en Lectura, Matemática, y Ciencias Sociales

4.3.1 Calibración de los ítems

En las tablas 4.1 a la 4.7, se presenta los bloques, la posición por bloques, medida de dificultad, indicadores de ajuste, tasa de acierto (p) y nivel de desempeño para cada uno de los ítems que conforman las distintas pruebas aplicadas en la ENLA muestral 2023. En el caso de los ítems de créditos parciales, se presenta los resultados para cada umbral. Las tablas 4.1, 4.2, 4.3 y 4.4 presentan las medidas para los ítems de las pruebas de 2.º y 4.º grados de primaria, y las tablas 4.5, 4.6 y 4.7 contienen información para los ítems de las pruebas de 2.º grado de secundaria.

Tabla 4.1 Dificultad y ajuste de los ítems al modelo Rasch, Lectura – 2.º grado de primaria

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT001	B13	1	0,760	0,077	0,78	0,70	0,59	0,68	3
CT002	B13	3	1,021	0,075	0,80	0,73	0,59	0,62	3
CT003	B13	4	1,793	0,074	0,95	0,93	0,48	0,47	3
CT004	B26	2	1,126	0,074	0,96	0,91	0,47	0,60	3
CT005	B26	3	2,290	0,076	1,12	1,20	0,32	0,36	4
CT006	B07	4	1,429	0,023	1,01	1,09	0,41	0,67	3
CT007	B07	2	1,768	0,023	1,02	1,05	0,42	0,60	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT008	B26	4	2,331	0,076	1,04	1,17	0,37	0,35	4
CT009	B07	3	1,383	0,024	0,87	0,78	0,52	0,68	3
CT010	B26	1	0,875	0,076	1,01	0,92	0,42	0,65	3
CT011	B06	1	0,822	0,026	0,93	0,89	0,44	0,78	3
CT012	B06	2	2,721	0,023	1,15	1,25	0,31	0,41	4
CT013	B06	3	2,374	0,022	0,96	0,95	0,48	0,48	4
CT014	B06	4	1,976	0,022	1,14	1,21	0,32	0,56	3
CT015	B20	2	1,904	0,022	1,14	1,19	0,33	0,58	3
CT016	B20	1	0,471	0,028	0,84	0,66	0,50	0,82	3
CT017	B20	3	2,348	0,022	1,21	1,31	0,27	0,48	4
CT018	B20	4	2,140	0,022	1,01	1,03	0,43	0,53	4
CT019	B05	1	0,215	0,030	0,85	0,62	0,47	0,86	3
CT020	B05	2	0,203	0,030	0,84	0,64	0,47	0,86	3
CT021	B05	3	-0,337	0,036	0,83	0,51	0,44	0,91	2
CT022	B13	2	0,805	0,077	0,79	0,68	0,59	0,67	3
CT023	B21	5	2,360	0,031	1,00	1,02	0,44	0,48	4
CT024	B06	5	1,291	0,024	1,00	1,10	0,40	0,70	3
CT025	B07	1	0,222	0,030	0,83	0,61	0,48	0,86	3
CT026	B08	2	2,462	0,031	1,08	1,15	0,37	0,46	4
CT027	B08	3	3,023	0,033	1,09	1,20	0,35	0,35	4
CT028	B08	5	1,979	0,031	1,01	1,00	0,43	0,56	3
CT029	B08	4	2,455	0,031	1,23	1,33	0,25	0,46	4
CT030	B09	1	2,787	0,032	1,13	1,24	0,33	0,40	4
CT031	B09	2	1,966	0,032	1,00	0,99	0,44	0,56	3
CT032	B09	4	2,163	0,031	0,94	0,90	0,50	0,52	4
CT033	B06	6	1,908	0,022	0,95	0,93	0,48	0,58	3
CT034	B08	1	1,661	0,032	0,99	0,96	0,44	0,63	3
CT035	B22	1	-	-	-	-	-	-	-
CT036	B22	2	2,341	0,031	0,98	0,98	0,46	0,49	4
CT037	B22	3	2,515	0,031	1,25	1,40	0,24	0,45	4
CT038	B22	4	2,263	0,031	1,00	1,01	0,44	0,50	4
CT039	B22	5	2,067	0,031	1,08	1,13	0,38	0,54	4
CT040	B09	3	2,301	0,031	1,04	1,05	0,41	0,50	4
CT041	B09	5	2,553	0,032	1,12	1,15	0,35	0,44	4
CT042	B10	4	2,139	0,034	0,83	0,77	0,57	0,55	4
CT043	B11	1	2,219	0,035	0,86	0,81	0,55	0,54	4
CT044	B11	5	3,294	0,037	1,17	1,33	0,27	0,32	4

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT045	B11	3	3,642	0,039	1,05	1,25	0,34	0,26	4
CT046	B23	1	2,813	0,035	0,97	1,01	0,45	0,41	4
CT047	B23	2	3,328	0,037	0,96	1,08	0,43	0,31	4
CT048	B23	3	2,938	0,035	1,18	1,32	0,27	0,39	4
CT049	B23	4	-	-	-	-	-	-	-
CT050	B15	1	-0,768	0,112	0,96	0,83	0,35	0,89	2
CT051	B04	1	-0,185	0,049	0,92	0,85	0,37	0,90	2
CT052	B17	1	-0,099	0,047	0,95	0,90	0,35	0,89	2
CT053	B10	1	1,767	0,035	0,88	0,81	0,52	0,63	3
CT054	B25	1	1,734	0,073	1,00	1,01	0,42	0,49	3
CT055	B25	2	1,841	0,073	1,17	1,22	0,27	0,47	3
CT056	B25	3	2,204	0,074	1,23	1,40	0,20	0,39	4
CT057	B08	6	3,284	0,034	1,10	1,25	0,32	0,30	4
CT058	B05	4	0,474	0,028	0,89	0,81	0,45	0,82	3
CT059	B10	2	2,911	0,035	1,10	1,18	0,34	0,39	4
CT060	B10	3	2,307	0,034	1,00	1,00	0,44	0,52	4
CT061	B11	2	2,634	0,035	0,94	0,97	0,48	0,46	4
CT062	B11	4	2,734	0,035	1,00	1,02	0,43	0,43	4
CT063	B12	3	1,592	0,074	0,95	0,91	0,47	0,52	3
CT064	B18	4	0,820	0,026	0,97	1,00	0,41	0,78	3
CT065	B24	1	2,082	0,035	0,99	0,99	0,44	0,57	4
CT066	B24	2	1,645	0,036	0,93	0,88	0,47	0,66	3
CT067	B24	3	1,067	0,039	0,88	0,79	0,48	0,76	3
CT068	B24	4	1,644	0,036	0,96	0,97	0,44	0,66	3
CT069	B24	5	3,398	0,037	1,10	1,39	0,30	0,30	4
CT070	B18	1	0,134	0,031	0,86	0,67	0,45	0,86	3
CT071	B18	2	0,381	0,029	0,81	0,61	0,50	0,84	3
CT072	B18	3	0,038	0,032	0,87	0,68	0,44	0,87	2
CT073	B12	1	1,826	0,074	0,95	0,94	0,46	0,47	3
CT074	B12	2	2,099	0,075	0,99	0,99	0,42	0,42	4
CT075	B21	1	3,000	0,033	1,20	1,45	0,25	0,35	4
CT076	B21	2	2,472	0,031	1,04	1,11	0,40	0,46	4
CT077	B21	6	2,161	0,031	0,96	0,97	0,47	0,52	4
CT078	B21	3	2,549	0,031	1,06	1,09	0,39	0,44	4
CT079	B21	4	2,947	0,032	0,90	0,91	0,50	0,36	4
CT080	B19	1	1,685	0,023	0,89	0,83	0,52	0,62	3
CT081	B19	2	1,920	0,022	0,96	0,95	0,47	0,57	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT082	B19	3	2,578	0,022	1,02	1,08	0,42	0,44	4
CT083	B19	4	2,431	0,022	0,92	0,92	0,50	0,47	4
CT084	B19	5	3,111	0,023	1,09	1,23	0,35	0,33	4
CT085	B19	6	2,580	0,022	1,22	1,32	0,27	0,44	4
CT086	B02	1	-0,256	0,095	1,03	1,08	0,30	0,84	2
CT087	B16	1	-0,484	0,054	0,87	0,65	0,39	0,92	2
CT088	B14	1	-0,055	0,092	0,81	0,58	0,53	0,82	2
CT089	B03	1	-0,381	0,052	0,89	0,68	0,39	0,91	2
CT090	B01	1	-0,776	0,118	0,94	0,70	0,36	0,90	2

Tabla 4.2 Dificultad y ajuste de los ítems al modelo Rasch, Matemática – 2.º grado de primaria

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA001	B11	4	-0,626	0,036	1,06	1,27	0,38	0,73	2
MA002	B11	7	0,453	0,033	0,87	0,88	0,57	0,55	3
MA003	B12	6	1,136	0,034	0,84	0,85	0,61	0,41	3
MA004	B13	1	1,321	0,034	0,81	0,74	0,64	0,38	3
MA005	B12	2	0,854	0,033	0,88	0,84	0,59	0,46	3
MA006	B11	8	-0,354	0,034	0,97	1,04	0,46	0,69	2
MA007	B12	3	-0,097	0,034	1,16	1,40	0,38	0,63	2
MA008	B13	2	-0,913	0,037	0,96	0,99	0,45	0,77	2
MA009	B09	7	0,387	0,033	0,94	0,93	0,52	0,56	3
MA010	B14	3	0,902	0,033	0,94	0,95	0,54	0,46	3
MA011	B11	3	1,320	0,033	1,10	1,17	0,44	0,39	3
MA012	B10	8	-0,408	0,035	0,81	0,73	0,58	0,68	2
MA013	B12	5	-0,266	0,034	0,91	0,81	0,54	0,66	2
MA014	B11	1	-0,800	0,037	1,02	0,97	0,41	0,76	2
MA015	B08	5	-0,069	0,055	1,01	0,98	0,50	0,54	2
MA016	B10	3	0,173	0,033	0,96	1,06	0,52	0,58	3
MA017	B09	6	0,349	0,033	1,00	0,97	0,49	0,57	3
MA018	B12	7	0,420	0,033	0,99	0,99	0,52	0,54	3
MA019	B15	1	0,638	0,033	1,24	1,33	0,38	0,50	3
MA020	B16	1	0,889	0,033	1,03	1,06	0,48	0,46	3
MA021	B04	5	2,463	0,030	1,00	1,14	0,47	0,22	4
MA022	B02	6	0,699	0,025	1,01	1,01	0,50	0,52	3
MA023	B09	2	1,113	0,033	1,27	1,45	0,33	0,43	3
MA024	B03	2	2,245	0,029	1,12	1,40	0,40	0,25	4

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA025	B12	8	-0,087	0,034	1,12	1,24	0,41	0,63	2
MA026	B10	2	-	-	-	-	-	-	-
MA027	B13	3	1,873	0,036	1,13	1,47	0,41	0,30	4
MA028	B01	4	1,955	0,028	1,12	1,32	0,41	0,30	4
MA029	B11	6	1,684	0,035	1,13	1,32	0,41	0,33	3
MA030	B06	5	0,017	0,055	0,94	0,95	0,54	0,53	2
MA031	B10	6	-	-	-	-	-	-	-
MA032	B04	2	1,769	0,027	0,83	0,81	0,60	0,33	3
MA033	B05	5	0,356	0,055	0,89	0,90	0,58	0,47	3
MA034	B16	2	0,432	0,033	1,17	1,26	0,39	0,55	3
MA035	B09	5	1,194	0,033	1,13	1,22	0,42	0,41	3
MA036	B04	4	1,563	0,027	1,19	1,30	0,39	0,36	3
MA037	B08	1	-2,523	0,081	0,98	1,34	0,30	0,90	1
MA038	B10	1	-1,295	0,040	0,95	0,96	0,42	0,81	2
MA039	B09	1	-0,944	0,038	1,01	1,22	0,38	0,78	2
MA040	B10	5	0,671	0,033	0,91	0,90	0,57	0,49	3
MA041	B08	3	-1,206	0,060	1,07	1,20	0,38	0,74	2
MA042	B06	4	-1,215	0,060	0,99	0,93	0,44	0,74	2
MA043	B07	3	-0,246	0,055	0,79	0,73	0,62	0,58	2
MA044	B15	2	0,068	0,033	0,81	0,74	0,61	0,60	2
MA045	B01	2	2,175	0,029	1,10	1,31	0,42	0,26	4
MA046	B10	7	1,226	0,034	0,92	0,95	0,57	0,40	3
MA047	B07	6	-1,770	0,066	0,89	1,00	0,45	0,82	1
MA048	B12	1	-1,513	0,042	0,91	1,01	0,42	0,84	1
MA049	B02	2	2,405	0,030	1,04	1,19	0,45	0,23	4
MA050	B06	2	0,402	0,055	0,97	0,94	0,54	0,46	3
MA051	B06	3	-0,107	0,055	0,98	0,96	0,52	0,55	2
MA052	B09	3	1,853	0,035	0,92	0,92	0,54	0,30	3
MA053	B04	1	0,561	0,025	0,86	0,81	0,59	0,54	3
MA054	B03	5	2,216	0,029	1,06	1,11	0,45	0,26	4
MA055	B16	3	1,167	0,033	1,19	1,30	0,39	0,41	3
MA056	B03	1	-1,321	0,032	0,97	0,91	0,39	0,84	2
MA057	B02	4	0,899	0,026	1,09	1,18	0,45	0,48	3
MA058	B01	1	0,703	0,025	0,91	0,88	0,56	0,52	3
MA059	B03	6	-0,046	0,026	0,89	0,80	0,54	0,65	2
MA060	B06	1	-2,432	0,079	0,97	1,06	0,32	0,89	1
MA061	B01	3	1,712	0,027	0,99	0,99	0,51	0,34	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA062	B12	4	1,553	0,035	0,97	1,17	0,52	0,34	3
MA063	B11	5	0,305	0,033	1,03	1,07	0,47	0,57	3
MA064	B05	4	-0,412	0,055	1,18	1,16	0,38	0,61	2
MA065	B11	2	0,342	0,033	0,92	0,89	0,54	0,57	3
MA066	B05	3	0,114	0,055	1,06	1,06	0,47	0,51	3
MA067	B10	4	-0,215	0,034	0,98	1,05	0,49	0,65	2
MA068	B09	4	-0,246	0,034	0,83	0,73	0,57	0,67	2
MA069	B04	6	1,656	0,027	1,07	1,31	0,44	0,34	3
MA070	B01	6	1,048	0,026	1,28	1,45	0,33	0,45	3
MA071	B09	8	-1,319	0,041	0,98	1,10	0,38	0,83	2
MA072	B08	2	-0,447	0,055	0,70	0,62	0,67	0,61	2
MA073	B08	4	-0,454	0,055	0,87	0,81	0,57	0,61	2
MA074	B01	5	1,826	0,027	1,14	1,30	0,41	0,32	3
MA075	B02	3	1,640	0,027	0,96	1,00	0,53	0,35	3
MA076	B04	3	-0,574	0,028	0,78	0,64	0,57	0,74	2
MA077	B07	4	-0,124	0,055	1,20	1,48	0,36	0,55	2
MA078	B14	2	0,565	0,033	1,00	1,00	0,50	0,52	3
MA079	B07	5	-1,112	0,059	0,88	0,89	0,51	0,72	2
MA080	B05	2	0,484	0,055	0,97	0,99	0,53	0,44	3
MA081	B08	6	-0,661	0,056	0,89	0,85	0,54	0,65	2
MA082	B07	1	-1,711	0,065	0,84	0,71	0,49	0,81	1
MA083	B05	6	-1,261	0,060	0,91	0,88	0,48	0,75	2
MA084	B07	2	-1,494	0,063	0,98	0,92	0,42	0,78	1
MA085	B02	5	2,438	0,030	1,05	1,23	0,44	0,23	4
MA086	B03	4	0,373	0,026	0,87	0,85	0,57	0,57	3
MA087	B15	3	1,897	0,036	0,76	0,67	0,64	0,29	4
MA088	B03	3	0,976	0,026	0,87	0,84	0,59	0,46	3
MA089	B06	6	-2,213	0,075	0,94	0,88	0,37	0,87	1
MA090	B05	1	-3,104	0,100	0,95	0,93	0,28	0,94	1
MA091	B02	1	-0,304	0,027	0,96	1,01	0,47	0,69	2
MA092	B14	1	0,436	0,033	0,81	0,76	0,61	0,55	3

Los ítems CT035 y CT049 de la prueba de Lectura, y MA026 y MA031 de la prueba de Matemática no presentaron un buen ajuste al modelo Rasch. Por ello, no fueron incluidos en la estimación de las medidas de habilidad de los estudiantes evaluados en 2.º grado de primaria.

Tabla 4.3 Dificultad y ajuste de los ítems al modelo Rasch, Lectura – 4.º grado de primaria

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT001	B05	1	1,023	0,012	1,16	1,43	0,33	0,31	4
CT002	B05	2	-	-	-	-	-	-	-
CT003	B05	3	0,615	0,012	1,04	1,15	0,43	0,38	4
CT004	B05	4	-0,543	0,011	0,93	0,88	0,52	0,58	3
CT005	B05	5	-1,706	0,013	1,02	1,08	0,38	0,77	2
CT006	B15	2	-2,631	0,018	0,83	0,65	0,48	0,82	1
CT007	B15	4	-3,312	0,021	0,84	0,59	0,43	0,89	1
CT008	B11	5	-0,807	0,011	0,95	0,91	0,49	0,63	3
CT009	B09	2	-0,876	0,020	1,15	1,23	0,31	0,55	3
CT010	B09	3	-0,312	0,020	1,08	1,12	0,38	0,44	3
CT011	B09	5	-1,304	0,020	1,03	1,02	0,39	0,63	2
CT012	B09	4	-0,474	0,020	1,11	1,14	0,36	0,48	3
CT013	B09	1	-1,755	0,021	0,98	0,96	0,41	0,70	2
CT014	B09	6	-0,048	0,021	1,14	1,19	0,34	0,40	3
CT015	B15	6	-2,403	0,017	0,92	0,86	0,43	0,79	1
CT016	B10	5	0,002	0,021	1,25	1,35	0,26	0,38	3
CT017	B04	4	-0,480	0,006	1,02	1,01	0,45	0,58	3
CT018	B15	1	-3,020	0,020	0,87	0,69	0,43	0,86	1
CT019	B15	3	-2,973	0,019	0,83	0,59	0,46	0,86	1
CT020	B15	5	-2,277	0,016	0,84	0,73	0,49	0,78	1
CT021	B12	1	-0,087	0,009	1,01	1,02	0,43	0,56	3
CT022	B12	2	0,203	0,009	0,98	0,97	0,47	0,50	4
CT023	B12	3	-0,254	0,009	0,86	0,80	0,55	0,58	3
CT024	B12	5	1,186	0,010	1,16	1,38	0,29	0,33	4
CT025	B12	6	0,194	0,009	1,06	1,10	0,40	0,50	4
CT026	B06	6	0,650	0,009	1,05	1,08	0,41	0,42	4
CT027	B03	5	-1,134	0,007	0,88	0,79	0,52	0,69	2
CT028	B06	5	-1,223	0,010	0,99	1,00	0,40	0,75	2
CT029	B12	4	-1,173	0,010	0,83	0,72	0,53	0,74	2
CT030	B10	6	-0,959	0,020	0,95	0,92	0,48	0,55	2
CT031	B11	1	0,226	0,011	1,04	1,09	0,45	0,44	4
CT032	B11	2	-0,998	0,012	1,00	1,04	0,44	0,66	2
CT033	B11	6	-1,408	0,012	0,87	0,77	0,51	0,72	2
CT034	B11	3	-0,936	0,011	0,95	0,88	0,49	0,65	2
CT035	B11	4	-0,416	0,011	0,95	0,93	0,50	0,56	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT036	B14	1	-1,356	0,012	0,86	0,73	0,53	0,73	2
CT037	B14	2	-1,004	0,012	0,93	0,86	0,51	0,68	2
CT038	B14	3	-0,264	0,011	0,97	0,97	0,50	0,55	3
CT039	B14	4	-0,664	0,011	0,85	0,77	0,58	0,62	3
CT040	B14	5	0,283	0,011	1,08	1,14	0,42	0,46	4
CT041	B14	6	0,184	0,011	1,26	1,40	0,30	0,47	4
CT042	B08	1	-1,294	0,015	0,87	0,78	0,49	0,77	2
CT043	B08	5	-0,796	0,014	0,88	0,79	0,53	0,70	3
CT044	B08	2	-1,862	0,017	0,86	0,70	0,46	0,84	2
CT045	B08	4	-1,966	0,018	0,91	0,82	0,42	0,85	2
CT046	B16	2	-0,536	0,014	0,94	0,92	0,50	0,48	3
CT047	B16	3	-0,711	0,014	0,96	0,93	0,48	0,51	3
CT048	B16	4	-0,003	0,015	1,01	1,06	0,44	0,38	3
CT049	B16	5	0,041	0,015	1,08	1,13	0,39	0,38	3
CT050	B01	1	-0,909	0,007	0,93	0,88	0,50	0,66	3
CT051	B01	2	-1,404	0,007	0,92	0,82	0,48	0,74	2
CT052	B01	4	-0,453	0,007	0,94	0,91	0,51	0,58	3
CT053	B01	6	0,459	0,082	0,77	0,69	0,64	0,41	4
CT054	B08	3	0,646	0,013	1,04	1,08	0,43	0,44	4
CT055	B08	6	-1,608	0,016	0,86	0,74	0,48	0,81	2
CT056	B01	5	0,552	0,007	1,20	1,28	0,33	0,40	4
CT057	B04	5	-	-	-	-	-	-	-
CT058	B01	3	-0,962	0,007	1,01	0,99	0,44	0,67	2
CT059	B13	1	0,756	0,009	0,95	0,95	0,49	0,41	4
CT060	B13	2	1,379	0,010	1,16	1,34	0,29	0,31	4
CT061	B13	3	0,513	0,009	1,15	1,20	0,33	0,46	4
CT062	B13	4	0,038	0,009	1,04	1,04	0,41	0,55	3
CT063	B13	5	0,728	0,009	1,05	1,11	0,40	0,42	4
CT064	B13	6	0,582	0,009	1,05	1,10	0,40	0,44	4
CT065	B02	1	-0,845	0,007	0,85	0,77	0,56	0,65	3
CT066	B02	2	0,535	0,007	1,19	1,29	0,33	0,40	4
CT067	B02	3	0,022	0,006	1,01	1,00	0,47	0,49	3
CT068	B02	4	-	-	-	-	-	-	-
CT069	B02	5	-0,688	0,007	0,87	0,79	0,55	0,62	3
CT070	B02	6	1,994	0,102	0,90	0,74	0,48	0,18	4
CT071	B07	2	-0,413	0,010	1,03	1,00	0,41	0,63	3
CT072	B07	3	1,262	0,010	1,14	1,37	0,30	0,33	4

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT073	B07	4	0,076	0,009	1,03	1,06	0,42	0,54	3
CT074	B04	6	2,438	0,113	0,95	0,90	0,40	0,13	4
CT075	B05	6	0,566	0,011	1,14	1,24	0,37	0,39	4
CT076	B07	1	-0,269	0,009	1,04	1,02	0,41	0,60	3
CT077	B07	5	0,511	0,009	1,08	1,11	0,39	0,46	4
CT078	B07	6	0,136	0,009	1,13	1,18	0,34	0,53	3
CT079	B16	1	0,755	0,016	1,18	1,38	0,30	0,26	4
CT080	B16	6	-0,191	0,014	0,99	1,02	0,46	0,42	3
CT081	B03	4	-1,607	0,007	0,93	0,94	0,45	0,76	2
CT082	B03	3	-0,404	0,006	0,87	0,82	0,56	0,57	3
CT083_1	B03	6	-0,117	0,058	0,97	0,90	0,62	0,72	3
CT083_2	B03	6	1,532	0,058	0,97	0,90	0,62	0,72	4
CT084	B06	1	0,416	0,009	1,06	1,09	0,40	0,46	4
CT085	B06	2	-1,096	0,010	0,88	0,81	0,49	0,73	2
CT086	B06	3	0,740	0,010	1,16	1,23	0,32	0,40	4
CT087	B06	4	-1,268	0,011	0,97	0,97	0,41	0,75	2
CT088	B10	1	-1,041	0,020	0,96	0,95	0,46	0,57	2
CT089	B10	2	0,015	0,021	0,92	0,90	0,52	0,38	3
CT090	B10	3	-0,705	0,020	1,09	1,11	0,37	0,50	3
CT091	B10	4	-0,480	0,020	1,11	1,14	0,36	0,46	3
CT092	B04	1	0,719	0,007	1,06	1,12	0,42	0,37	4
CT093	B04	2	0,384	0,006	1,08	1,15	0,41	0,43	4
CT094	B04	3	-0,130	0,006	1,05	1,06	0,44	0,52	3
CT095	B03	1	-1,046	0,007	0,85	0,77	0,55	0,68	2
CT096	B03	2	-0,554	0,007	0,96	0,94	0,49	0,60	3

Tabla 4.4 Dificultad y ajuste de los ítems al modelo Rasch, Matemática – 4.º grado de primaria

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA001	B18	1	-1,582	0,013	1,01	1,00	0,46	0,72	2
MA002	B18	6	-2,051	0,013	0,86	0,72	0,52	0,79	1
MA003	B06	2	-	-	-	-	-	-	-
MA004	B16	1	-1,701	0,013	0,83	0,75	0,56	0,74	2
MA005	B16	6	-1,550	0,012	0,94	0,86	0,50	0,72	2
MA006	B07	2	-2,702	0,014	1,03	1,25	0,37	0,79	1
MA007	B08	2	-1,629	0,010	0,96	0,93	0,50	0,69	2
MA008	B05	3	-1,151	0,012	1,00	0,97	0,50	0,55	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA009	B07	7	-2,411	0,013	0,89	0,80	0,50	0,75	1
MA010	B08	7	-2,715	0,011	0,83	0,67	0,50	0,83	1
MA011	B13	5	-2,130	0,014	0,89	0,81	0,49	0,80	1
MA012	B15	1	-1,907	0,013	0,89	0,82	0,50	0,77	2
MA013	B14	6	-1,436	0,012	0,89	0,80	0,54	0,70	2
MA014	B04	7	-1,935	0,010	0,82	0,62	0,53	0,81	2
MA015	B07	4	-1,353	0,012	1,10	1,13	0,43	0,58	2
MA016	B15	4	0,472	0,012	1,04	1,16	0,47	0,40	4
MA017	B02	4	0,416	0,008	1,28	1,46	0,28	0,44	4
MA018	B07	5	-1,815	0,012	0,99	0,94	0,48	0,66	2
MA019	B13	6	-0,788	0,012	1,09	1,17	0,42	0,60	3
MA020	B13	4	0,153	0,011	1,25	1,48	0,32	0,45	3
MA021	B17	1	-2,771	0,016	0,99	1,19	0,38	0,86	1
MA022	B15	2	-1,743	0,013	0,95	0,96	0,48	0,75	2
MA023	B13	2	-0,572	0,011	0,89	0,87	0,56	0,57	3
MA024	B13	3	0,556	0,012	1,06	1,14	0,44	0,38	4
MA025	B17	2	-0,827	0,012	0,98	0,99	0,51	0,60	3
MA026	B05	2	-1,412	0,012	1,02	1,07	0,47	0,60	2
MA027	B03	1	-2,539	0,013	1,00	1,04	0,34	0,87	1
MA028	B15	3	1,201	0,013	1,13	1,43	0,38	0,28	4
MA029	B14	4	0,397	0,012	1,06	1,24	0,44	0,40	4
MA030	B18	3	0,641	0,012	1,14	1,34	0,40	0,37	4
MA031	B17	3	-1,745	0,013	0,89	0,84	0,52	0,74	2
MA032	B16	4	-0,232	0,011	1,05	1,09	0,46	0,51	3
MA033	B02	5	1,461	0,009	1,01	1,10	0,44	0,28	4
MA034	B06	6	-0,653	0,009	1,05	1,07	0,48	0,54	3
MA035	B03	3	-	-	-	-	-	-	-
MA036	B01	3	1,359	0,010	1,19	1,50	0,30	0,29	4
MA037	B04	4	2,320	0,010	1,01	1,43	0,36	0,17	4
MA038	B15	6	-1,619	0,013	0,96	0,96	0,48	0,73	2
MA039	B07	1	-2,731	0,014	0,90	0,86	0,46	0,79	1
MA040	B17	4	0,713	0,012	1,19	1,37	0,36	0,35	4
MA041	B05	5	-1,523	0,012	0,95	0,90	0,51	0,62	2
MA042	B14	5	-1,086	0,012	0,97	0,97	0,51	0,65	2
MA043	B06	5	0,072	0,009	1,22	1,40	0,37	0,42	3
MA044	B04	3	1,316	0,008	1,08	1,31	0,38	0,30	4
MA045	B03	2	-0,386	0,010	0,96	0,96	0,50	0,58	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA046	B06	7	-1,735	0,010	0,88	0,83	0,54	0,71	2
MA047	B01	1	-1,493	0,011	0,78	0,64	0,58	0,76	2
MA048	B08	3	-0,417	0,009	0,88	0,86	0,59	0,50	3
MA049	B16	5	-0,283	0,011	1,04	1,08	0,47	0,52	3
MA050	B17	5	-0,116	0,011	0,98	1,00	0,52	0,49	3
MA051	B02	3	-0,498	0,008	1,19	1,23	0,34	0,60	3
MA052	B14	2	-1,292	0,012	0,94	0,91	0,52	0,68	2
MA053	B01	4	0,794	0,010	1,21	1,43	0,31	0,38	4
MA054	B01	2	0,420	0,010	1,04	1,09	0,45	0,44	4
MA055	B03	5	-	-	-	-	-	-	-
MA056	B02	6	0,091	0,008	1,01	1,02	0,47	0,50	3
MA057	B16	2	-0,753	0,012	0,93	0,94	0,53	0,60	3
MA058	B04	5	-0,857	0,008	1,10	1,11	0,40	0,66	3
MA059	B18	4	0,930	0,012	1,11	1,44	0,41	0,32	4
MA060	B05	1	-3,384	0,016	0,96	1,19	0,35	0,86	1
MA061	B04	2	0,082	0,008	0,99	1,00	0,49	0,50	3
MA062	B05	4	-0,894	0,012	1,17	1,23	0,40	0,51	3
MA063	B08	6	-0,882	0,009	1,04	1,05	0,48	0,57	3
MA064	B06	4	-	-	-	-	-	-	-
MA065	B18	2	-0,882	0,012	0,96	0,93	0,52	0,62	3
MA066	B03	4	0,542	0,010	1,07	1,14	0,42	0,42	4
MA067	B05	6	-1,377	0,012	0,82	0,73	0,60	0,59	2
MA068	B11	1	0,614	0,085	0,82	0,74	0,61	0,36	4
MA069	B01	7	-1,451	0,011	0,96	0,90	0,46	0,75	2
MA070	B03	6	0,404	0,010	1,06	1,12	0,43	0,44	4
MA071	B02	2	-0,543	0,008	0,88	0,81	0,56	0,61	3
MA072	B06	1	-3,154	0,013	0,91	0,76	0,42	0,87	1
MA073	B05	7	-2,946	0,015	0,98	1,05	0,38	0,82	1
MA074	B14	3	0,447	0,012	1,10	1,25	0,42	0,39	4
MA075	B03	7	-1,072	0,010	1,01	1,05	0,44	0,70	2
MA076	B06	3	-0,248	0,009	1,14	1,22	0,43	0,47	3
MA077	B02	1	-1,088	0,008	0,84	0,76	0,56	0,70	2
MA078	B19	2	-0,307	0,007	0,92	0,90	0,56	0,52	3
MA079	B19	1	-0,286	0,007	1,08	1,12	0,45	0,52	3
MA080	B19	3	-0,473	0,007	0,86	0,82	0,59	0,55	3
MA081	B20	2	-0,364	0,007	0,97	0,99	0,52	0,53	3
MA082	B20	1	-1,546	0,007	0,87	0,81	0,54	0,72	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA083	B20	3	-1,475	0,007	1,03	1,04	0,45	0,71	2
MA084	B02	7	-0,300	0,008	1,04	1,03	0,45	0,57	3
MA085	B04	1	-2,448	0,011	0,87	0,77	0,44	0,86	1
MA086	B10	1	0,317	0,083	0,84	0,74	0,60	0,41	3
MA087_1	B12	1	-1,401	0,056	1,00	1,00	0,63	1,25	2
MA087_2	B12	1	-0,318	0,056	1,00	1,00	0,63	1,25	3
MA088	B09	1	1,773	0,100	1,00	0,90	0,44	0,20	4
MA089	B01	5	-	-	-	-	-	-	-
MA090	B08	4	-	-	-	-	-	-	-
MA091	B04	6	-	-	-	-	-	-	-
MA092	B01	6	-0,491	0,010	1,00	0,99	0,47	0,60	3
MA093	B08	5	-1,253	0,009	0,87	0,81	0,57	0,63	2
MA094	B15	5	-0,660	0,012	0,86	0,78	0,59	0,58	3
MA095	B16	3	-0,241	0,011	0,90	0,89	0,56	0,51	3
MA096	B08	1	-2,982	0,012	0,99	1,07	0,37	0,86	1
MA097	B14	1	-1,949	0,013	0,96	0,89	0,47	0,77	2
MA098	B18	5	-0,702	0,012	0,94	0,90	0,54	0,59	3
MA099	B13	1	-2,208	0,014	1,03	1,11	0,38	0,81	1
MA100	B07	6	-1,248	0,012	0,94	0,89	0,54	0,57	2
MA101	B17	6	-1,262	0,012	0,80	0,73	0,60	0,67	2
MA102	B07	3	-2,072	0,013	0,97	1,03	0,46	0,70	1

Los ítems CT002 CT057 y CT068 de la prueba de Lectura, así como MA003, MA035, MA055, MA064, MA089, MA090 y MA091 de la prueba de Matemática no presentaron un buen ajuste al modelo Rasch. Por ello, dichos ítems no fueron incluidos en la estimación de las medidas de habilidad de los estudiantes evaluados en 4.º grado de primaria.

Tabla 4.5 Dificultad y ajuste de los ítems al modelo Rasch, Lectura – 2.º grado de secundaria

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT001	B13	1	2,498	0,012	0,90	0,88	0,48	0,28	4
CT002	B13	2	1,758	0,011	0,99	1,01	0,42	0,44	3
CT003	B13	3	1,325	0,011	1,06	1,08	0,36	0,55	3
CT004_1	B13	4	0,351	0,008	1,13	1,14	0,46	1,21	2
CT004_2	B13	4	1,815	0,008	1,13	1,14	0,46	1,21	4
CT005_1	B13	5	0,573	0,008	1,15	1,18	0,44	1,06	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT005_2	B13	5	2,214	0,008	1,15	1,18	0,44	1,06	4
CT006	B12	2	1,604	0,025	1,04	1,09	0,35	0,37	3
CT007	B12	3	0,927	0,024	1,22	1,28	0,19	0,53	2
CT008	B12	4	0,328	0,025	0,84	0,77	0,53	0,67	2
CT009	B12	5	1,770	0,025	1,18	1,28	0,21	0,33	4
CT010	B19	6	0,145	0,018	0,92	0,85	0,45	0,71	2
CT011	B16	1	0,668	0,013	0,91	0,87	0,44	0,72	2
CT012	B16	2	1,196	0,012	1,09	1,13	0,31	0,61	3
CT013	B16	4	1,157	0,012	1,09	1,10	0,31	0,62	3
CT014	B16	3	1,309	0,012	0,99	0,98	0,41	0,58	3
CT015	B13	6	0,495	0,012	1,00	0,96	0,37	0,73	2
CT016	B10	6	-0,409	0,030	0,90	0,76	0,43	0,81	1
CT017	B16	5	1,183	0,012	0,84	0,79	0,54	0,61	3
CT018	B17	1	2,135	0,011	1,11	1,18	0,31	0,36	4
CT019	B17	2	2,769	0,012	0,96	1,11	0,40	0,23	4
CT020	B17	3	1,798	0,011	0,98	1,00	0,43	0,44	4
CT021	B17	4	2,022	0,011	0,85	0,86	0,54	0,38	4
CT022	B17	5	2,366	0,011	1,04	1,11	0,36	0,31	4
CT023	B17	6	2,101	0,011	0,96	0,95	0,45	0,37	4
CT024	B07	2	2,305	0,012	1,17	1,31	0,23	0,35	4
CT025	B07	3	1,243	0,012	1,06	1,05	0,34	0,60	3
CT026	B07	4	1,857	0,012	1,02	1,04	0,39	0,45	4
CT027	B03	2	0,662	0,007	0,94	0,89	0,45	0,69	2
CT028	B03	3	1,702	0,006	1,21	1,31	0,23	0,45	3
CT029	B03	4	0,910	0,007	1,05	1,11	0,35	0,63	2
CT030	B03	5	1,265	0,006	1,01	1,01	0,41	0,55	3
CT031	B03	7	0,827	0,081	1,03	1,02	0,39	0,65	2
CT032	B11	1	-0,333	0,029	0,95	1,02	0,38	0,80	1
CT033	B11	2	-1,364	0,042	0,88	0,80	0,37	0,92	1
CT034	B11	3	-0,028	0,027	0,88	0,78	0,48	0,74	2
CT035	B11	4	0,802	0,024	1,06	1,15	0,36	0,56	2
CT036	B11	5	0,108	0,026	0,93	0,85	0,45	0,72	2
CT037	B04	6	2,309	0,084	0,90	0,80	0,51	0,31	4
CT038	B01	1	-0,081	0,008	1,01	1,22	0,30	0,82	2
CT039	B01	2	-0,115	0,008	0,92	0,83	0,40	0,83	1
CT040	B01	3	1,312	0,006	0,92	0,89	0,49	0,54	3
CT041	B01	7	3,291	0,108	1,04	0,88	0,32	0,14	4

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT042	B14	1	-0,834	0,021	1,01	0,95	0,21	0,93	1
CT043	B14	3	2,081	0,012	0,99	0,99	0,42	0,39	4
CT044	B14	4	1,361	0,012	0,98	0,97	0,41	0,57	3
CT045	B14	5	2,209	0,012	1,07	1,16	0,32	0,36	4
CT046	B01	5	0,116	0,008	0,87	0,76	0,47	0,79	2
CT047	B03	6	0,706	0,007	1,08	1,15	0,31	0,68	2
CT048	B05	3	-0,162	0,018	0,86	0,73	0,45	0,83	1
CT049	B05	5	1,328	0,014	1,02	1,02	0,40	0,53	3
CT050	B08	2	1,114	0,017	0,98	0,95	0,41	0,63	3
CT051	B08	4	1,722	0,017	1,16	1,21	0,26	0,48	3
CT052	B08	5	0,759	0,018	0,92	0,87	0,44	0,70	2
CT053	B02	1	0,946	0,007	0,98	0,98	0,42	0,63	3
CT054	B02	3	1,499	0,006	1,01	1,02	0,41	0,50	3
CT055_1	B02	6	2,603	0,070	0,95	1,00	0,49	0,29	4
CT055_2	B02	7	3,402	0,070	0,95	1,00	0,49	0,29	4
CT056	B20	2	-0,288	0,020	0,90	0,80	0,43	0,79	1
CT057	B20	5	0,193	0,018	0,99	0,99	0,38	0,70	2
CT058	B18	1	0,420	0,018	0,98	0,96	0,41	0,64	2
CT059	B18	2	0,379	0,018	1,04	1,04	0,36	0,65	2
CT060	B18	3	0,270	0,018	0,93	0,86	0,46	0,68	2
CT061	B18	5	0,600	0,017	0,98	0,95	0,42	0,60	2
CT062	B18	4	0,570	0,017	1,01	1,00	0,39	0,61	2
CT063	B06	1	0,223	0,020	0,95	0,86	0,38	0,80	2
CT064	B03	1	0,755	0,007	0,95	0,91	0,44	0,67	2
CT065	B01	4	0,596	0,007	0,90	0,86	0,47	0,70	2
CT066	B01	6	1,248	0,006	1,08	1,12	0,34	0,56	3
CT067	B07	5	0,894	0,012	1,08	1,19	0,29	0,68	2
CT068	B07	6	0,924	0,012	1,03	1,05	0,34	0,67	2
CT069	B20	1	-0,316	0,020	0,99	1,16	0,32	0,80	1
CT070	B20	3	-0,427	0,021	0,91	0,80	0,41	0,81	1
CT071	B02	4	1,242	0,006	1,04	1,04	0,38	0,56	3
CT072	B05	1	0,532	0,015	0,93	0,89	0,44	0,71	2
CT073	B05	6	2,312	0,015	1,00	1,11	0,39	0,30	4
CT074	B14	2	1,763	0,012	1,07	1,10	0,33	0,47	3
CT075	B16	6	0,517	0,013	0,93	0,86	0,42	0,75	2
CT076	B11	6	0,366	0,025	0,94	0,92	0,45	0,66	2
CT077	B07	1	0,817	0,013	0,97	0,92	0,41	0,69	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT078	B02	2	1,370	0,006	0,98	0,97	0,44	0,53	3
CT079	B02	5	1,977	0,007	1,02	1,06	0,40	0,38	4
CT080	B04	2	0,916	0,007	0,90	0,86	0,49	0,63	2
CT081	B04	4	0,901	0,007	0,97	0,93	0,44	0,64	2
CT082	B05	2	0,426	0,015	1,03	1,03	0,35	0,73	2
CT083	B05	4	1,889	0,014	1,06	1,09	0,36	0,40	4
CT084	B06	6	1,549	0,017	1,00	1,00	0,41	0,52	3
CT085	B08	1	-	-	-	-	-	-	-
CT086	B08	3	1,383	0,017	1,14	1,19	0,27	0,56	3
CT087	B08	6	3,185	0,021	1,13	1,45	0,21	0,18	4
CT088	B12	1	0,490	0,025	1,09	1,11	0,30	0,64	2
CT089	B12	6	1,039	0,024	1,10	1,12	0,30	0,50	3
CT090	B14	6	0,862	0,012	0,96	0,92	0,41	0,68	2
CT091	B18	6	1,186	0,017	0,87	0,85	0,53	0,46	3
CT092	B19	4	1,716	0,018	1,07	1,12	0,35	0,34	3
CT093	B20	4	1,267	0,017	1,13	1,17	0,28	0,45	3
CT094	B20	6	1,299	0,017	0,93	0,93	0,46	0,44	3
CT095	B19	1	-1,222	0,028	0,92	0,71	0,35	0,91	1
CT096	B19	2	0,050	0,019	0,99	0,94	0,38	0,73	2
CT097	B19	3	1,029	0,017	0,94	0,93	0,47	0,50	3
CT098	B19	5	1,284	0,017	0,98	0,98	0,44	0,44	3
CT099	B15	1	1,483	0,012	1,06	1,08	0,34	0,54	3
CT100	B15	2	1,155	0,012	0,93	0,89	0,46	0,62	3
CT101	B15	3	1,507	0,012	1,02	1,02	0,39	0,53	3
CT102	B15	4	1,617	0,012	1,13	1,17	0,28	0,50	3
CT103	B15	5	1,686	0,012	0,97	0,98	0,43	0,49	3
CT104	B15	6	1,991	0,012	0,95	0,95	0,45	0,42	4
CT105	B06	2	1,191	0,017	1,05	1,06	0,34	0,61	3
CT106	B06	3	0,991	0,017	0,97	0,94	0,42	0,65	3
CT107	B06	4	0,230	0,020	0,95	0,82	0,39	0,80	2
CT108	B06	5	1,962	0,017	1,07	1,10	0,34	0,42	4
CT109	B04	1	-0,315	0,009	0,95	0,81	0,37	0,86	1
CT110	B04	3	0,859	0,007	1,06	1,10	0,34	0,64	2
CT111	B04	5	0,996	0,006	0,90	0,86	0,50	0,61	3
CT112	B10	1	0,783	0,024	0,88	0,84	0,51	0,57	2
CT113	B10	2	0,257	0,026	0,93	0,88	0,44	0,69	2
CT114	B10	3	0,539	0,025	1,11	1,17	0,29	0,62	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT115	B10	4	0,284	0,026	0,94	0,90	0,43	0,68	2
CT116	B10	5	1,249	0,024	1,00	1,02	0,40	0,45	3

Tabla 4.6 Dificultad y ajuste de los ítems al modelo Rasch, Matemática – 2.º grado de secundaria

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA001	B18	5	1,132	0,014	0,98	1,00	0,43	0,53	2
MA002	B18	3	2,241	0,015	0,92	1,11	0,45	0,30	4
MA003	B16	8	0,531	0,014	0,87	0,80	0,52	0,65	2
MA004	B17	8	0,726	0,011	0,98	0,99	0,41	0,64	2
MA005	B17	2	0,679	0,011	0,98	0,94	0,41	0,65	2
MA006	B06	2	0,401	0,007	1,11	1,20	0,28	0,69	2
MA007	B20	4	3,117	0,018	0,94	1,40	0,37	0,17	4
MA008	B19	4	1,660	0,014	0,89	0,92	0,50	0,42	3
MA009	B20	3	0,998	0,014	1,07	1,09	0,36	0,56	2
MA010	B03	3	1,565	0,012	1,16	1,24	0,25	0,48	3
MA011	B20	1	0,755	0,014	0,87	0,83	0,52	0,61	2
MA012	B14	1	-0,881	0,019	1,00	1,08	0,31	0,86	1
MA013	B18	6	1,299	0,014	0,97	0,99	0,44	0,49	2
MA014	B15	6	1,146	0,014	0,87	0,84	0,52	0,54	2
MA015	B13	5	2,397	0,011	0,88	0,90	0,49	0,30	4
MA016	B04	6	1,632	0,010	0,89	0,87	0,50	0,47	3
MA017	B19	3	1,724	0,014	0,96	1,01	0,44	0,41	3
MA018	B20	2	0,937	0,014	0,90	0,88	0,50	0,57	2
MA019	B03	4	2,898	0,014	0,94	1,01	0,42	0,23	4
MA020	B16	7	2,278	0,015	0,90	0,97	0,48	0,30	4
MA021	B06	4	0,821	0,006	0,97	0,96	0,43	0,61	2
MA022	B04	2	2,590	0,011	0,92	0,92	0,45	0,28	4
MA023	B13	3	2,526	0,012	0,92	0,97	0,46	0,28	4
MA024	B14	6	1,870	0,014	1,00	1,06	0,41	0,38	3
MA025	B15	5	1,854	0,014	1,19	1,30	0,23	0,39	3
MA026	B15	4	2,748	0,016	0,91	1,04	0,43	0,23	4
MA027	B01	5	0,097	0,014	1,00	0,98	0,34	0,77	1
MA028	B14	4	1,784	0,014	0,96	0,99	0,45	0,39	3
MA029	B03	1	0,320	0,013	1,00	1,00	0,36	0,74	2
MA030	B13	6	0,576	0,011	0,92	0,88	0,46	0,67	2
MA031	B15	2	1,944	0,014	0,95	0,98	0,44	0,37	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA032	B15	1	-0,434	0,017	0,92	0,83	0,42	0,82	1
MA033	B16	5	1,637	0,014	1,09	1,21	0,33	0,42	3
MA034	B08	4	0,191	0,007	0,94	0,90	0,44	0,72	1
MA035	B18	2	0,845	0,014	0,98	1,01	0,43	0,58	2
MA036	B16	6	1,140	0,014	0,90	0,90	0,50	0,53	2
MA037	B01	2	1,482	0,012	0,88	0,85	0,52	0,50	3
MA038	B20	5	1,969	0,014	1,06	1,20	0,35	0,36	3
MA039	B08	6	2,036	0,007	1,00	1,07	0,40	0,35	4
MA040	B04	3	2,787	0,011	1,04	1,20	0,31	0,24	4
MA041	B13	4	1,554	0,011	0,98	1,00	0,43	0,47	3
MA042	B06	3	-	-	-	-	-	-	-
MA043	B01	4	2,599	0,013	1,05	1,17	0,33	0,28	4
MA044	B20	6	1,911	0,014	1,01	1,11	0,40	0,37	3
MA045	B16	4	2,056	0,014	0,97	1,04	0,43	0,34	4
MA046	B03	5	2,609	0,013	1,07	1,16	0,32	0,28	4
MA047	B04	4	-	-	-	-	-	-	-
MA048	B04	1	0,549	0,010	0,91	0,88	0,44	0,69	2
MA049	B01	3	3,138	0,015	1,07	1,46	0,26	0,20	4
MA050	B02	3	2,362	0,010	0,97	0,97	0,42	0,32	4
MA051	B04	5	1,974	0,010	1,18	1,24	0,23	0,40	3
MA052	B02	1	1,066	0,010	1,11	1,19	0,27	0,59	2
MA053	B02	6	2,907	0,011	1,11	1,30	0,24	0,23	4
MA054	B21	3	2,007	0,007	1,18	1,37	0,23	0,36	3
MA055	B21	2	2,241	0,007	1,14	1,27	0,27	0,32	4
MA056	B19	5	1,226	0,014	0,98	1,00	0,43	0,52	2
MA057	B14	5	1,522	0,014	1,03	1,09	0,38	0,45	3
MA058	B02	2	2,478	0,010	1,02	1,20	0,34	0,30	4
MA059	B08	5	2,262	0,007	1,11	1,26	0,29	0,31	4
MA060	B21	1	-	-	-	-	-	-	-
MA061	B10	1	2,983	0,097	0,98	0,89	0,40	0,20	4
MA062	B11	1	3,844	0,124	0,96	0,80	0,35	0,10	4
MA063	B02	4	2,599	0,011	1,09	1,23	0,27	0,28	4
MA064	B01	6	-	-	-	-	-	-	-
MA065	B17	5	2,166	0,011	1,10	1,26	0,30	0,34	4
MA066	B02	5	1,634	0,010	0,99	0,99	0,41	0,47	3
MA067	B03	6	-	-	-	-	-	-	-
MA068	B06	1	-0,584	0,008	0,98	1,08	0,34	0,84	1

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA069	B19	1	0,159	0,015	0,99	0,97	0,39	0,73	1
MA070	B12	1	3,689	0,118	0,95	0,78	0,37	0,11	4
MA071	B22	3	1,838	0,006	0,97	1,07	0,42	0,39	3
MA072	B09	1	2,828	0,094	0,95	0,86	0,44	0,22	4
MA073	B07	4	-2,175	0,023	0,87	0,66	0,40	0,93	1
MA074	B07	1	-2,288	0,024	0,88	0,67	0,38	0,94	1
MA075	B05	3	-1,844	0,021	0,89	0,80	0,39	0,91	1
MA076	B13	2	0,701	0,011	0,94	0,89	0,45	0,64	2
MA077	B19	8	-0,256	0,017	0,88	0,78	0,47	0,79	1
MA078	B20	7	0,318	0,015	0,96	0,93	0,44	0,69	2
MA079	B13	1	-0,201	0,013	0,84	0,71	0,49	0,80	1
MA080	B01	1	0,351	0,013	0,87	0,79	0,48	0,73	2
MA081	B16	1	0,036	0,015	0,98	1,00	0,40	0,74	1
MA082	B14	2	1,025	0,014	1,00	1,00	0,41	0,55	2
MA083	B18	8	0,408	0,014	0,89	0,84	0,50	0,67	2
MA084	B17	3	1,102	0,011	1,00	1,01	0,40	0,56	2
MA085	B05	1	-1,078	0,016	0,86	0,81	0,47	0,84	1
MA086	B07	5	-1,014	0,016	0,88	0,80	0,46	0,83	1
MA087	B05	4	-1,351	0,018	0,87	0,85	0,44	0,87	1
MA088	B19	2	0,499	0,014	0,98	0,98	0,41	0,67	2
MA089	B13	8	-0,290	0,013	0,89	0,77	0,44	0,81	1
MA090	B07	6	-1,318	0,017	0,82	0,60	0,51	0,86	1
MA091	B05	2	-0,623	0,014	0,88	0,78	0,49	0,78	1
MA092	B15	3	1,747	0,014	0,94	0,93	0,46	0,41	3
MA093	B08	2	1,174	0,006	1,08	1,17	0,34	0,53	2
MA094	B22	1	1,047	0,006	1,00	1,00	0,42	0,56	2
MA095	B22	2	2,728	0,007	1,13	1,46	0,24	0,23	4
MA096	B08	3	-0,607	0,008	1,09	1,27	0,24	0,84	1
MA097	B18	7	1,261	0,014	0,92	0,92	0,49	0,50	2
MA098	B06	6	0,866	0,006	0,96	0,94	0,45	0,60	2
MA099	B15	7	0,891	0,014	1,08	1,14	0,33	0,59	2
MA100	B20	8	-0,166	0,016	0,83	0,70	0,52	0,77	1
MA101	B05	6	-1,317	0,017	0,94	0,92	0,39	0,87	1
MA102	B16	2	0,733	0,014	0,96	0,94	0,44	0,61	2
MA103	B15	8	0,693	0,014	0,86	0,80	0,53	0,63	2
MA104	B06	5	0,237	0,007	0,89	0,84	0,48	0,72	1
MA105	B18	4	1,063	0,014	0,94	0,93	0,47	0,54	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA106	B18	1	-0,708	0,018	1,01	0,92	0,34	0,84	1
MA107	B17	1	-1,060	0,017	0,91	0,78	0,37	0,90	1
MA108	B14	8	0,643	0,014	0,90	0,87	0,49	0,63	2
MA109	B07	3	-1,421	0,018	0,86	0,66	0,47	0,88	1
MA110	B08	1	-0,658	0,009	0,88	0,80	0,43	0,85	1
MA111	B17	6	1,355	0,011	0,91	0,88	0,49	0,51	3
MA112	B07	2	-1,105	0,016	0,95	0,92	0,40	0,84	1
MA113	B14	3	1,464	0,014	0,97	0,99	0,44	0,46	3
MA114	B19	6	1,782	0,014	0,96	1,02	0,43	0,40	3
MA115	B19	7	0,690	0,014	0,93	0,91	0,46	0,63	2
MA116	B17	4	1,310	0,011	1,08	1,10	0,34	0,52	3
MA117	B16	3	1,946	0,014	1,15	1,31	0,28	0,36	3
MA118	B14	7	0,885	0,014	1,05	1,08	0,37	0,58	2
MA119	B05	5	-0,205	0,013	0,91	0,82	0,49	0,71	1
MA120	B13	7	1,738	0,011	1,16	1,27	0,26	0,43	3
MA121	B03	2	2,289	0,012	1,08	1,17	0,32	0,34	4
MA122	B17	7	0,391	0,011	0,88	0,80	0,50	0,70	2

Tabla 4.7 Dificultad y ajuste de los ítems al modelo Rasch, Ciencias Sociales – 2.º grado de secundaria

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CS001	B03	6	-1,842	0,022	0,86	0,67	0,45	0,86	1
CS002	B02	6	-0,950	0,017	0,88	0,83	0,48	0,74	2
CS003	B07	2	-1,091	0,007	0,96	0,99	0,39	0,76	1
CS004	B03	5	0,587	0,015	0,95	0,96	0,43	0,46	3
CS005	B21	4	-0,540	0,021	1,00	1,00	0,40	0,59	2
CS006	B05	6	1,135	0,016	0,99	1,00	0,38	0,35	4
CS007	B02	5	0,143	0,015	0,99	0,99	0,41	0,54	3
CS008	B07	1	0,083	0,006	0,92	0,90	0,48	0,56	3
CS009	B01	6	-1,593	0,020	0,88	0,77	0,45	0,83	1
CS010	B12	1	-0,544	0,014	0,96	0,93	0,41	0,70	2
CS011	B04	5	0,735	0,016	0,96	0,97	0,45	0,43	3
CS012	B09	4	-0,865	0,013	1,01	1,04	0,38	0,65	2
CS013	B04	6	0,504	0,016	0,95	0,94	0,46	0,47	3
CS014	B06	5	-0,764	0,017	0,91	0,86	0,46	0,70	2
CS015	B06	6	-0,536	0,016	0,90	0,86	0,49	0,66	2
CS016	B11	6	-0,692	0,014	0,84	0,75	0,53	0,73	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CS017	B14	6	0,505	0,013	0,99	0,99	0,38	0,50	3
CS018	B16	2	-0,183	0,021	1,10	1,11	0,33	0,53	3
CS019	B14	5	1,088	0,013	1,00	1,03	0,35	0,39	4
CS020	B15	6	0,531	0,013	1,03	1,04	0,35	0,49	3
CS021	B17	5	-0,087	0,021	0,92	0,89	0,49	0,51	3
CS022	B19	5	-1,487	0,024	0,86	0,76	0,49	0,76	1
CS023	B11	5	0,986	0,013	0,99	1,00	0,42	0,42	4
CS024	B20	6	0,315	0,021	1,02	1,03	0,38	0,43	3
CS025	B17	4	-0,759	0,022	0,91	0,86	0,49	0,64	2
CS026	B19	6	-0,047	0,021	0,96	0,95	0,45	0,50	3
CS027	B16	1	-0,900	0,022	0,87	0,80	0,52	0,66	2
CS028	B10	1	0,562	0,013	0,86	0,83	0,54	0,49	3
CS029	B13	1	0,232	0,013	1,02	1,02	0,35	0,56	3
CS030_1	B07	3	-1,141	0,039	1,08	1,08	0,48	1,30	1
CS030_2	B07	3	0,207	0,039	1,08	1,08	0,48	1,30	3
CS031	B09	5	-1,064	0,013	0,99	0,98	0,40	0,69	1
CS032	B18	4	-0,341	0,021	1,04	1,07	0,38	0,55	2
CS033	B15	3	2,173	0,016	1,11	1,45	0,16	0,20	4
CS034	B05	4	-0,453	0,016	0,93	0,89	0,45	0,66	2
CS035	B07	5	-0,713	0,007	0,85	0,76	0,52	0,70	2
CS036	B04	3	-1,113	0,018	0,95	0,90	0,42	0,76	1
CS037	B08	4	0,159	0,007	0,93	0,90	0,46	0,57	3
CS038	B06	3	-1,184	0,018	0,91	0,84	0,44	0,77	1
CS039	B07	4	-0,336	0,007	0,98	0,97	0,42	0,64	2
CS040	B20	3	-1,828	0,026	0,87	0,74	0,47	0,81	1
CS041	B04	4	-0,687	0,017	0,97	0,93	0,43	0,70	2
CS042	B03	4	-0,751	0,017	1,03	1,03	0,33	0,72	2
CS043	B02	4	0,543	0,015	1,00	1,02	0,39	0,46	3
CS044	B06	4	-0,629	0,016	0,98	0,96	0,41	0,68	2
CS045	B01	4	0,428	0,016	1,02	1,04	0,41	0,50	3
CS046	B01	3	-1,348	0,019	0,90	0,81	0,45	0,80	1
CS047	B05	5	-0,898	0,017	0,98	0,95	0,39	0,74	2
CS048	B08	5	-0,386	0,008	0,95	0,91	0,43	0,68	2
CS049	B10	4	0,963	0,013	0,98	1,02	0,41	0,41	4
CS050	B08	6	-0,276	0,008	0,97	0,96	0,41	0,66	2
CS051	B21	3	-0,413	0,021	0,90	0,86	0,50	0,57	2
CS052	B15	2	0,099	0,013	1,00	1,00	0,38	0,58	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CS053	B10	6	-0,291	0,013	0,99	0,97	0,40	0,65	2
CS054	B13	3	1,064	0,013	1,10	1,14	0,26	0,39	4
CS055	B21	2	-1,217	0,023	0,94	0,91	0,43	0,71	1
CS056	B12	6	0,583	0,013	1,01	1,03	0,38	0,49	3
CS057	B13	2	0,143	0,013	0,92	0,89	0,46	0,58	3
CS058	B16	3	-0,452	0,021	1,02	1,03	0,39	0,58	2
CS059	B12	5	0,642	0,013	1,09	1,11	0,30	0,48	3
CS060	B16	4	-0,511	0,021	0,87	0,83	0,53	0,59	2
CS061	B14	4	-1,258	0,016	0,90	0,81	0,41	0,82	1
CS062	B20	5	-0,284	0,021	0,92	0,90	0,48	0,54	2
CS063	B11	4	0,503	0,013	0,99	1,00	0,42	0,51	3
CS064	B15	1	0,043	0,013	1,12	1,19	0,24	0,59	3
CS065	B19	3	0,104	0,021	0,98	0,99	0,42	0,47	3
CS066	B18	6	-0,733	0,022	0,93	0,89	0,47	0,63	2
CS067	B19	1	-0,211	0,021	1,00	1,00	0,41	0,53	3
CS068	B01	5	0,446	0,016	1,01	1,02	0,41	0,49	3
CS069	B14	3	1,133	0,013	1,03	1,07	0,32	0,38	4
CS070	B18	5	-0,734	0,022	0,96	0,95	0,44	0,63	2
CS071	B17	6	-0,283	0,021	0,87	0,83	0,53	0,54	2
CS072	B13	4	1,895	0,015	1,12	1,40	0,16	0,24	4
CS073	B10	5	-0,516	0,014	0,94	0,90	0,44	0,69	2
CS074	B21	1	0,371	0,021	1,04	1,07	0,35	0,41	3
CS075_1	B09	3	-0,738	0,051	1,03	1,05	0,53	1,07	2
CS075_2	B09	3	0,322	0,051	1,03	1,05	0,53	1,07	3
CS076	B19	2	-0,832	0,022	0,96	0,91	0,44	0,65	2
CS077	B20	4	0,596	0,022	1,11	1,18	0,28	0,37	3
CS078	B13	5	0,983	0,013	1,03	1,05	0,34	0,41	4
CS079	B06	2	0,154	0,016	0,93	0,92	0,47	0,53	3
CS080	B11	2	1,683	0,014	1,11	1,20	0,28	0,29	4
CS081	B03	2	1,285	0,016	1,04	1,14	0,30	0,32	4
CS082	B03	1	0,801	0,016	1,14	1,21	0,23	0,42	3
CS083	B21	6	-0,271	0,021	1,11	1,15	0,30	0,54	2
CS084	B11	3	1,177	0,013	1,02	1,09	0,37	0,38	4
CS085	B04	2	0,706	0,016	1,10	1,16	0,32	0,44	3
CS086	B05	3	-1,278	0,019	0,85	0,70	0,50	0,79	1
CS087	B20	2	1,338	0,024	1,10	1,32	0,24	0,25	4
CS088	B17	2	0,570	0,022	1,01	1,02	0,41	0,38	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CS089	B06	1	0,144	0,016	1,07	1,10	0,34	0,53	3
CS090	B03	3	-0,312	0,016	0,92	0,87	0,47	0,64	2
CS091	B16	5	0,270	0,021	1,07	1,09	0,36	0,44	3
CS092	B04	1	0,435	0,016	1,13	1,17	0,30	0,49	3
CS093	B02	3	0,762	0,016	1,03	1,06	0,36	0,42	3
CS094	B02	1	-0,468	0,016	0,94	0,91	0,44	0,66	2
CS095	B01	2	-0,296	0,016	0,97	0,96	0,44	0,63	2
CS096	B13	6	0,894	0,013	1,09	1,12	0,27	0,42	4
CS097	B01	1	1,158	0,016	1,17	1,30	0,25	0,36	4
CS098	B02	2	-0,459	0,016	1,01	1,00	0,38	0,66	2
CS099	B07	6	0,265	0,006	1,07	1,10	0,33	0,52	3
CS100	B05	2	1,326	0,016	1,07	1,18	0,28	0,32	4
CS101	B05	1	-0,817	0,017	1,03	1,03	0,34	0,72	2
CS102	B10	2	1,120	0,013	1,10	1,18	0,30	0,38	4
CS103	B17	1	0,447	0,022	1,19	1,26	0,26	0,40	3
CS104	B09	1	-1,101	0,013	0,86	0,77	0,52	0,70	1
CS105	B09	6	-0,998	0,013	1,10	1,16	0,29	0,68	2
CS106	B15	4	1,588	0,014	1,01	1,09	0,33	0,29	4
CS107	B12	4	1,186	0,013	1,12	1,20	0,25	0,37	4
CS108	B16	6	1,492	0,025	1,21	1,47	0,16	0,23	4
CS109	B18	1	1,236	0,024	1,21	1,45	0,19	0,27	4
CS110	B19	4	-0,546	0,021	1,04	1,04	0,37	0,60	2
CS111	B15	5	0,396	0,013	1,05	1,06	0,33	0,52	3
CS112	B14	2	2,138	0,015	1,14	1,44	0,12	0,21	4
CS113	B20	1	-0,072	0,021	1,14	1,18	0,27	0,50	3
CS114	B09	2	-0,242	0,012	1,00	1,00	0,41	0,54	3
CS115	B12	3	0,291	0,013	0,96	0,95	0,43	0,55	3
CS116	B08	2	1,028	0,007	1,17	1,24	0,21	0,40	4
CS117	B10	3	0,548	0,013	0,97	0,97	0,43	0,49	3
CS118	B08	1	0,968	0,007	1,00	1,02	0,38	0,41	4
CS119_1	B08	3	-1,217	0,057	1,01	1,01	0,47	1,33	1
CS119_2	B08	3	0,549	0,057	1,01	1,01	0,47	1,33	3
CS120	B11	1	-0,163	0,013	1,00	1,01	0,39	0,64	3
CS121	B12	2	-0,106	0,013	1,00	1,02	0,38	0,62	3
CS122	B14	1	0,968	0,013	1,09	1,13	0,26	0,41	4
CS123	B17	3	-0,319	0,021	0,90	0,86	0,51	0,55	2
CS124	B18	2	-1,474	0,024	0,97	0,95	0,39	0,75	1

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CS125	B18	3	-	-	-	-	-	-	-
CS126	B21	5	0,652	0,022	1,19	1,33	0,20	0,36	3

Los ítems CT002 CT057 y CT068 de la prueba de Lectura, así como MA003, MA035, MA055, MA064, MA089, MA090 y MA091 de la prueba de Matemática no presentaron un buen ajuste al modelo Rasch. Por ello, dichos ítems no fueron incluidos en la estimación de las medidas de habilidad de los estudiantes evaluados en 4.º grado de primaria.

4.3.2 Evidencias de validez vinculadas a la estructura interna

Para la ENLA muestral 2023, como se precisó con anterioridad, luego de ajustar el modelo Rasch unidimensional se realizó un análisis de componentes principales de los residuos estandarizados. Los resultados se presentan en la tabla 4.8.

Tabla 4.8 Análisis de unidimensionalidad de las medidas derivadas de las pruebas de la ENLA muestral 2023

Prueba	Varianza de la dimensión principal (%)	Varianza modelada (%)	Primer autovalor	Primer autovalor (%)
Lectura, 2.º grado de primaria	29,7	29,9	2,61	2,1
Matemática, 2.º grado de primaria	35,5	34,9	1,97	1,4
Lectura, 4.º grado de primaria	27,2	27,2	2,08	1,6
Matemática, 4.º grado de primaria	32,5	32,2	1,76	1,2
Lectura, 2.º grado de secundaria	24,1	24,3	1,43	0,9
Matemática, 2.º grado de secundaria	31,1	30,8	1,77	1,0
Ciencia Sociales, 2.º grado de secundaria	22,9	22,7	1,50	0,9

Linacre (2024) afirma que, aunque no existen parámetros absolutos para interpretar los resultados del análisis de componentes principales de los residuos, si el primer autovalor contiene menos del 5 % de varianza o si su valor es menor que 3, entonces no existen evidencias serias que atenten contra el supuesto de unidimensionalidad; por tanto, los datos se pueden analizar adecuadamente utilizando modelos Rasch.

Además, se está frente a datos unidimensionales cuando la cantidad de varianza empírica es similar a la cantidad de varianza predicha por el modelo. Todas estas condiciones se cumplen en las pruebas analizadas. Por ende, no hay evidencias contundentes en contra del supuesto de unidimensionalidad del conjunto de ítems incluidos en estos instrumentos de medición.

4.3.3 Confiabilidad y consistencia de la clasificación

Para expresar los resultados de estos análisis, se utilizan tanto el índice de confiabilidad de la separación de las personas (Rp) como el índice de separación de personas (Gp).

Tabla 4.9 Análisis de confiabilidad de las medidas derivadas de la aplicación de las pruebas de la ENLA muestral 2023

Prueba	Rp	Gp
Lectura, 2.º grado de primaria	0,89	2,87
Matemática, 2.º grado de primaria	0,91	3,23
Lectura, 4.º grado de primaria	0,89	2,79
Matemática, 4.º grado de primaria	0,90	3,03
Lectura, 2.º grado de secundaria	0,86	2,45
Matemática, 2.º grado de secundaria	0,87	2,56
Ciencia Sociales, 2.º grado de secundaria	0,80	1,97

Considerando el valor de Rp para la aplicación de la ENLA muestral 2023, se observa que la varianza de error se encuentra entre 9 % y 24 % para las pruebas de primaria y secundaria. Por lo tanto, es posible afirmar que las medidas derivadas de aplicar dichas pruebas poseen adecuadas evidencias de confiabilidad.

Como se señaló previamente, otros indicadores importantes son los índices de precisión y consistencia de clasificación, cuyos resultados se presentan en las tablas 4.10, 4.11 y 4.12.

Tabla 4.10 Indicadores del análisis de precisión y consistencia de la clasificación usando las medidas derivadas de las pruebas de 2.º grado de primaria

	Lectura		Matemática	
	Precisión	Consistencia	Precisión	Consistencia
Nacional	0,87	0,82	0,83	0,76
Hombre	0,88	0,83	0,83	0,76
Mujer	0,87	0,82	0,82	0,75
Urbano	0,88	0,83	0,83	0,76
Rural	0,86	0,80	0,82	0,75
Estatad	0,87	0,82	0,83	0,76
No estadad	0,88	0,83	0,83	0,76
Polidocente Completo	0,88	0,83	0,83	0,76
Unidocente / Multigrado	0,86	0,80	0,83	0,75
Alta habilidad	0,88	0,83	0,83	0,76
Baja habilidad	0,86	0,80	0,82	0,75

Tabla 4.11 Indicadores del análisis de precisión y consistencia de la clasificación usando las medidas derivadas de las pruebas de 4.º grado de primaria

	Lectura		Matemática	
	Precisión	Consistencia	Precisión	Consistencia
Nacional	0,79	0,71	0,81	0,73
Hombre	0,79	0,71	0,81	0,74
Mujer	0,80	0,71	0,80	0,73
Urbano	0,80	0,72	0,81	0,73
Rural	0,77	0,68	0,82	0,74
Estatad	0,79	0,71	0,81	0,73
No estadad	0,81	0,74	0,80	0,72
Polidocente Completo	0,80	0,72	0,81	0,73
Unidocente / Multigrado	0,78	0,69	0,82	0,75
Alta habilidad	0,81	0,73	0,80	0,73
Baja habilidad	0,77	0,68	0,82	0,74

Tabla 4.12 Indicadores del análisis de precisión y consistencia de la clasificación usando las medidas derivadas de las pruebas de 2.º grado de secundaria

	Lectura		Matemática		Ciencias Sociales	
	Precisión	Consistencia	Precisión	Consistencia	Precisión	Consistencia
Nacional	0,72	0,62	0,75	0,65	0,71	0,60
Hombre	0,72	0,62	0,75	0,65	0,71	0,61
Mujer	0,72	0,62	0,75	0,66	0,70	0,60
Urbano	0,72	0,62	0,74	0,65	0,70	0,60
Rural	0,73	0,63	0,77	0,69	0,72	0,62
Estatad	0,72	0,62	0,75	0,66	0,70	0,60
No estadad	0,72	0,62	0,73	0,63	0,71	0,61
Alta habilidad	0,72	0,62	0,74	0,64	0,70	0,60
Baja habilidad	0,73	0,63	0,77	0,69	0,72	0,62

4.4 Equiparación de medidas de Educación Básica Regular (EBR)

En el marco del análisis Rasch, las puntuaciones del rasgo latente tienen un origen y una escala de medición arbitrarios (Bond et al., 2021). Por ello, dos instrumentos que miden lo mismo, aplicados a distintas muestras de personas, no se encontrarán en la misma métrica. Para establecer las comparaciones necesarias, ambas pruebas deben equipararse; es decir, deben ser puestas en la misma escala. La conversión implica usar una constante aditiva y, en otras, el uso de una constante multiplicativa (Livingston, 2004).

Diversas síntesis teóricas que tratan sobre la equiparación de puntuaciones (Kolen y Brennan, 2014; Navas, 1996), citan los trabajos de Angoff y Lord (ambos de la primera mitad de la década de 1980) como los pioneros. Estos autores entienden la equiparación como el desarrollo de un sistema de conversión de las unidades de una prueba a las unidades de otra, de manera que sus resultados sean comparables o equivalentes.

La finalidad es tener una métrica común para dos o más medidas de un mismo rasgo, de modo que se puedan comparar los resultados de personas a las que se les aplican diferentes instrumentos que miden el mismo rasgo. Por ejemplo, si se deseara equiparar dos pruebas distintas X e Y, el objetivo sería encontrar un sistema o función para convertir la métrica de X en Y.

Considerando la literatura sobre el tema (Ho y Osborn, 2005; Kolen y Brennan, 2014; Livingston, 2004; Navas, 1996; Zhu, 1998), se puede señalar que existen tres tipos de diseños de equiparación: de un solo grupo, de grupos equivalentes y de grupos no equivalentes con ítems comunes. Este último diseño fue utilizado para colocar las medidas de las pruebas de las distintas evaluaciones de 2.º grado de primaria en la misma métrica que las pruebas aplicadas en el 2007 por la UMC. En este diseño, se trabaja con dos muestras de personas que no necesariamente han sido extraídas de la misma población; en cada muestra, se aplica una única forma de la prueba. Lo esencial de este método es que, en cada grupo, se administra un conjunto de ítems o una prueba común que permite establecer la equivalencia entre las pruebas que se quieren equiparar.

Una vez obtenidos los datos, se utiliza la transformación lineal: se consideran equivalentes las puntuaciones directas que corresponden a la misma puntuación típica. En el contexto de la teoría de respuesta al ítem (TRI), se denomina a este método mean/sigma (Kolen y Brennan, 2014), e implica el cálculo de dos constantes de equiparación a y b, de tal manera que surge la siguiente fórmula:

$$medida_{equiparada} = b + a \times medida_{actual}$$

Donde:

$$a = \frac{S_{medida_{anterior}}}{S_{medida_{actual}}}$$
$$b = \bar{X}_{medida_{anterior}} - a \times \bar{X}_{medida_{anterior}}$$

Un requisito importante para realizar la comparación del rendimiento entre dos o más grupos es asegurar la invarianza de la dificultad de los ítems de la prueba a lo largo del tiempo (Kolen y Brennan, 2004). Esto supone que el grado de dificultad de cada ítem debe mantenerse constante², independientemente del grupo poblacional en el que se aplique. En caso de no cumplirse con este supuesto, se dice que el ítem presenta un funcionamiento diferencial (FDI) a lo largo del tiempo. Por ello, todos los ítems utilizados para el proceso de equiparación son sometidos a un análisis de FDI según la metodología propuesta por Bond et al. (2021). De esta manera, solo se utilizan en la equiparación los ítems sin FDI a lo largo del tiempo.

Desde la tabla 4.15 hasta la 4.23, se presentan los resultados del proceso de equiparación por ítems comunes de las pruebas aplicadas en la ENLA muestral 2023. Dichos resultados nos permiten señalar que este proceso de equiparación ha sido realizado siguiendo rigurosos criterios técnicos, que sustentan la comparación entre los resultados de las evaluaciones realizadas en años diferentes.

Tabla 4.13 Análisis de FDI de las medidas, prueba de Lectura de 2.º grado de primaria, usadas en la equiparación de 2022 con 2023

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT006	1,385	0,006	-0,315	0,023	1,429	-0,044	0,024
CT007	1,717	0,006	0,041	0,023	1,768	-0,051	0,024
CT009	1,252	0,006	-0,364	0,024	1,383	-0,131	0,025
CT011	0,759	0,007	-0,954	0,026	0,822	-0,063	0,027
CT012	2,732	0,006	1,043	0,023	2,720	0,012	0,024
CT013	2,118	0,006	0,678	0,022	2,373	-0,255	0,023
CT014	1,839	0,006	0,260	0,022	1,976	-0,137	0,023
CT015	1,819	0,006	0,184	0,022	1,904	-0,085	0,023
CT016	0,589	0,007	-1,323	0,028	0,471	0,118	0,029
CT017	2,217	0,006	0,651	0,022	2,348	-0,131	0,023
CT018	2,165	0,006	0,432	0,022	2,139	0,026	0,023
CT023	2,376	0,008	0,664	0,031	2,360	0,016	0,032
CT024	1,166	0,006	-0,461	0,024	1,290	-0,124	0,025
CT025	0,202	0,007	-1,585	0,030	0,222	-0,020	0,031
CT026	2,566	0,008	0,771	0,031	2,462	0,104	0,032
CT027	2,971	0,009	1,361	0,033	3,023	-0,052	0,034
CT028	2,070	0,008	0,263	0,031	1,979	0,091	0,032
CT029	2,449	0,008	0,763	0,031	2,454	-0,005	0,032

² En realidad, la dificultad del ítem debe mantenerse dentro de ciertos márgenes de error al comparar su dificultad en dos poblaciones diferentes.

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT030	2,839	0,009	1,113	0,032	2,787	0,052	0,033
CT031	1,999	0,008	0,249	0,032	1,965	0,034	0,033
CT032	2,095	0,008	0,456	0,031	2,162	-0,067	0,032
CT033	1,980	0,006	0,188	0,022	1,907	0,073	0,023
CT034	1,721	0,008	-0,071	0,032	1,661	0,060	0,033
CT036	2,360	0,008	0,644	0,031	2,341	0,019	0,032
CT037	2,601	0,008	0,827	0,031	2,515	0,086	0,032
CT038	2,205	0,008	0,561	0,031	2,262	-0,057	0,032
CT039	1,816	0,008	0,355	0,031	2,066	-0,250	0,032
CT040	2,357	0,008	0,601	0,031	2,300	0,057	0,032
CT041	2,546	0,009	0,866	0,032	2,552	-0,006	0,033
CT051	0,141	0,010	-2,013	0,049	-0,185	0,326	0,050
CT052	-0,108	0,011	-1,922	0,047	-0,099	-0,009	0,048
CT057	3,427	0,009	1,635	0,034	3,283	0,144	0,035
CT075	3,008	0,009	1,337	0,033	3,000	0,008	0,034
CT076	2,555	0,008	0,781	0,031	2,471	0,084	0,032
CT077	2,179	0,008	0,454	0,031	2,160	0,019	0,032
CT078	2,580	0,008	0,862	0,031	2,548	0,032	0,032
CT079	3,056	0,009	1,281	0,032	2,947	0,109	0,033
CT080	1,703	0,006	-0,046	0,023	1,685	0,018	0,024
CT081	1,870	0,006	0,201	0,022	1,920	-0,050	0,023
CT082	2,631	0,006	0,893	0,022	2,578	0,053	0,023
CT083	2,377	0,006	0,738	0,022	2,430	-0,053	0,023
CT084	3,216	0,006	1,453	0,023	3,110	0,106	0,024
CT085	2,509	0,006	0,895	0,022	2,580	-0,071	0,023
CT087	-0,565	0,013	-2,327	0,054	-0,484	-0,081	0,056
CT089	-0,286	0,012	-2,219	0,052	-0,381	0,095	0,053

Tabla 4.14 Análisis de FDI de las medidas, prueba de Matemática de 2.º grado de primaria, usadas en la equiparación de 2022 con 2023

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA001	-0,655	0,009	-0,898	0,036	-0,625	-0,030	0,037

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA002	0,619	0,009	0,187	0,033	0,453	0,166	0,034
MA003	0,949	0,009	0,874	0,034	1,136	-0,187	0,035
MA005	1,001	0,009	0,591	0,033	0,855	0,146	0,034
MA006	-0,292	0,009	-0,625	0,034	-0,354	0,062	0,035
MA007	0,176	0,011	-0,366	0,034	-0,096	0,272	0,036
MA008	-0,662	0,009	-1,187	0,037	-0,912	0,250	0,038
MA009	0,386	0,009	0,121	0,033	0,388	-0,002	0,034
MA010	1,023	0,007	0,639	0,033	0,902	0,121	0,034
MA011	1,330	0,009	1,059	0,033	1,320	0,010	0,034
MA012	-0,791	0,009	-0,679	0,035	-0,408	-0,383	0,036
MA013	-0,122	0,011	-0,536	0,034	-0,265	0,143	0,036
MA014	-0,878	0,009	-1,073	0,037	-0,799	-0,079	0,038
MA016	0,449	0,009	-0,095	0,033	0,173	0,276	0,034
MA017	0,161	0,009	0,082	0,033	0,349	-0,188	0,034
MA018	0,337	0,011	0,154	0,033	0,420	-0,083	0,035
MA019	0,649	0,007	0,373	0,033	0,638	0,011	0,034
MA023	0,782	0,009	0,851	0,033	1,113	-0,331	0,034
MA025	-0,120	0,009	-0,356	0,034	-0,087	-0,033	0,035
MA027	1,751	0,010	1,616	0,036	1,873	-0,122	0,037
MA029	1,640	0,009	1,426	0,035	1,685	-0,045	0,036
MA034	0,513	0,009	0,166	0,033	0,432	0,081	0,034
MA035	1,028	0,009	0,933	0,033	1,195	-0,167	0,034
MA038	-1,368	0,010	-1,571	0,040	-1,294	-0,074	0,041
MA039	-1,269	0,010	-1,218	0,038	-0,943	-0,326	0,039
MA040	0,614	0,009	0,406	0,033	0,671	-0,057	0,034
MA044	0,084	0,009	-0,200	0,033	0,069	0,015	0,034
MA046	1,282	0,009	0,965	0,034	1,226	0,056	0,035
MA048	-1,168	0,010	-1,791	0,042	-1,513	0,345	0,043
MA052	2,026	0,010	1,596	0,035	1,854	0,172	0,036
MA055	0,862	0,007	0,905	0,033	1,167	-0,305	0,034
MA062	1,507	0,008	1,294	0,035	1,553	-0,046	0,036
MA063	0,501	0,009	0,038	0,033	0,305	0,196	0,034
MA065	0,447	0,011	0,075	0,033	0,342	0,105	0,035
MA067	-0,096	0,009	-0,485	0,034	-0,215	0,119	0,035
MA068	0,069	0,011	-0,516	0,034	-0,246	0,315	0,036

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA071	-1,644	0,012	-1,596	0,041	-1,319	-0,325	0,043
MA078	0,486	0,009	0,300	0,033	0,565	-0,079	0,034

Tabla 4.15 Análisis de FDI de las medidas, prueba de Lectura de 4.º grado de primaria, usadas en la equiparación de 2022 con 2023

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT001	1,123	0,012	1,346	0,012	1,023	0,1	0,017
CT003	0,962	0,012	0,962	0,012	0,613	0,349	0,017
CT004	-0,258	0,011	-0,123	0,011	-0,544	0,286	0,016
CT008	-0,835	0,012	-0,37	0,011	-0,808	-0,027	0,016
CT017	-0,455	0,006	-0,063	0,006	-0,48	0,025	0,008
CT027	-1,06	0,007	-0,676	0,007	-1,134	0,074	0,01
CT031	0,091	0,011	0,599	0,011	0,226	-0,135	0,016
CT032	-1,091	0,012	-0,549	0,012	-0,999	-0,092	0,017
CT033	-1,39	0,013	-0,933	0,012	-1,408	0,018	0,018
CT034	-0,9	0,012	-0,49	0,011	-0,936	0,036	0,016
CT035	-0,456	0,011	-0,003	0,011	-0,416	-0,04	0,016
CT036	-1,392	0,013	-0,885	0,012	-1,357	-0,035	0,018
CT037	-0,813	0,011	-0,555	0,012	-1,005	0,192	0,016
CT038	-0,278	0,011	0,139	0,011	-0,265	-0,013	0,016
CT039	-0,592	0,011	-0,235	0,011	-0,664	0,072	0,016
CT040	0,446	0,011	0,652	0,011	0,283	0,163	0,016
CT041	-0,059	0,011	0,559	0,011	0,183	-0,242	0,016
CT043	-1,212	0,012	-0,359	0,014	-0,796	-0,416	0,018
CT050	-0,904	0,007	-0,466	0,007	-0,91	0,006	0,01
CT051	-1,208	0,007	-0,929	0,007	-1,404	0,196	0,01
CT052	-0,435	0,006	-0,039	0,007	-0,455	0,02	0,009
CT053	0,384	0,081	0,817	0,082	0,459	-0,075	0,115
CT054	0,394	0,011	0,992	0,013	0,645	-0,251	0,017
CT056	0,45	0,006	0,904	0,007	0,551	-0,101	0,009
CT058	-0,894	0,007	-0,515	0,007	-0,962	0,068	0,01
CT074	2,613	0,128	2,672	0,113	2,438	0,175	0,171
CT075	0,37	0,011	0,916	0,011	0,564	-0,194	0,016

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT081	-1,492	0,007	-1,119	0,007	-1,607	0,115	0,01
CT082	-0,396	0,006	0,007	0,006	-0,405	0,009	0,008
CT092	0,586	0,007	1,06	0,007	0,718	-0,132	0,01
CT093	0,258	0,006	0,747	0,006	0,384	-0,126	0,008
CT094	-0,159	0,006	0,265	0,006	-0,13	-0,029	0,008
CT095	-0,993	0,007	-0,593	0,007	-1,046	0,053	0,01
CT096	-0,601	0,007	-0,132	0,007	-0,554	-0,047	0,01

Tabla 4.16 Análisis de FDI de las medidas, prueba de Matemática de 4.º grado de primaria, usadas en la equiparación de 2022 con 2023

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA001	-1,559	0,012	-0,685	0,013	-1,581	0,022	0,018
MA002	-2,429	0,014	-1,118	0,013	-2,051	-0,378	0,019
MA004	-1,523	0,011	-0,795	0,013	-1,700	0,177	0,017
MA005	-1,774	0,013	-0,656	0,012	-1,550	-0,224	0,018
MA011	-2,175	0,010	-1,191	0,014	-2,130	-0,045	0,017
MA012	-2,209	0,013	-0,985	0,013	-1,906	-0,303	0,018
MA013	-1,573	0,012	-0,551	0,012	-1,436	-0,137	0,017
MA016	0,435	0,009	1,209	0,012	0,472	-0,037	0,015
MA019	-0,839	0,011	0,047	0,012	-0,788	-0,051	0,016
MA020	-0,225	0,011	0,915	0,011	0,153	-0,378	0,016
MA021	-2,498	0,015	-1,782	0,016	-2,770	0,272	0,022
MA022	-1,902	0,013	-0,834	0,013	-1,743	-0,159	0,018
MA023	-0,321	0,011	0,246	0,011	-0,572	0,251	0,016
MA024	0,469	0,011	1,287	0,012	0,557	-0,088	0,016
MA025	-0,704	0,012	0,011	0,012	-0,827	0,123	0,017
MA028	1,211	0,012	1,882	0,013	1,202	0,009	0,018
MA029	0,313	0,011	1,140	0,012	0,397	-0,084	0,016
MA030	0,374	0,011	1,365	0,012	0,641	-0,267	0,016
MA031	-1,736	0,013	-0,836	0,013	-1,745	0,009	0,018
MA032	0,065	0,011	0,560	0,011	-0,232	0,297	0,016
MA038	-1,818	0,013	-0,720	0,013	-1,619	-0,199	0,018
MA040	0,975	0,012	1,432	0,012	0,714	0,261	0,017

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA042	-1,037	0,012	-0,228	0,012	-1,086	0,049	0,017
MA049	-0,420	0,009	0,513	0,011	-0,282	-0,138	0,014
MA052	-0,967	0,012	-0,418	0,012	-1,292	0,325	0,017
MA057	-0,393	0,009	0,079	0,012	-0,753	0,360	0,015
MA059	0,724	0,012	1,632	0,012	0,931	-0,207	0,017
MA065	-0,661	0,010	-0,040	0,012	-0,882	0,221	0,016
MA074	0,366	0,011	1,186	0,012	0,447	-0,081	0,016
MA094	-0,639	0,011	0,165	0,012	-0,660	0,021	0,016
MA095	-0,190	0,011	0,552	0,011	-0,240	0,050	0,016
MA097	-1,769	0,013	-1,024	0,013	-1,949	0,180	0,018
MA098	-0,717	0,012	0,126	0,012	-0,702	-0,015	0,017
MA099	-2,149	0,013	-1,263	0,014	-2,208	0,059	0,019
MA101	-1,155	0,012	-0,390	0,012	-1,261	0,106	0,017

Tabla 4.17 Análisis de FDI de las medidas, prueba de Lectura de 2.º grado de secundaria, usadas en la equiparación de 2022 con 2023

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT001	2,230	0,011	1,719	0,012	2,497	-0,267	0,016
CT002	1,813	0,010	0,851	0,011	1,758	0,055	0,015
CT003	1,339	0,010	0,343	0,011	1,325	0,014	0,015
CT015	0,545	0,011	-0,632	0,012	0,495	0,050	0,016
CT016	-0,312	0,017	-1,692	0,030	-0,408	0,096	0,034
CT018	2,468	0,011	1,293	0,011	2,134	0,334	0,016
CT019	2,818	0,012	2,037	0,012	2,768	0,050	0,017
CT020	1,887	0,010	0,898	0,011	1,798	0,089	0,015
CT021	1,984	0,010	1,161	0,011	2,022	-0,038	0,015
CT022	2,331	0,011	1,564	0,011	2,365	-0,034	0,016
CT023	2,078	0,011	1,254	0,011	2,101	-0,023	0,016
CT027	0,643	0,006	-0,436	0,007	0,662	-0,019	0,009
CT028	1,585	0,006	0,785	0,006	1,702	-0,117	0,008
CT029	0,853	0,006	-0,144	0,007	0,911	-0,058	0,009
CT030	1,385	0,006	0,272	0,006	1,265	0,120	0,008
CT031	0,921	0,057	-0,242	0,081	0,827	0,094	0,099

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT038	-0,112	0,013	-1,308	0,008	-0,081	-0,031	0,015
CT039	-0,136	0,013	-1,348	0,008	-0,115	-0,021	0,015
CT040	1,535	0,010	0,327	0,006	1,312	0,223	0,012
CT046	0,366	0,011	-1,076	0,008	0,117	0,249	0,014
CT047	0,638	0,006	-0,384	0,007	0,706	-0,068	0,009
CT048	0,020	0,013	-1,403	0,018	-0,162	0,182	0,022
CT049	1,374	0,011	0,346	0,014	1,328	0,046	0,018
CT053	0,569	0,014	-0,102	0,007	0,946	-0,377	0,016
CT054	1,130	0,013	0,547	0,006	1,499	-0,369	0,014
CT064	0,715	0,006	-0,326	0,007	0,756	-0,041	0,009
CT065	0,691	0,011	-0,513	0,007	0,596	0,095	0,013
CT066	1,436	0,010	0,252	0,006	1,248	0,188	0,012
CT071	0,900	0,013	0,245	0,006	1,242	-0,342	0,014
CT072	0,809	0,011	-0,588	0,015	0,533	0,276	0,019
CT073	2,402	0,013	1,501	0,015	2,312	0,090	0,020
CT112	0,691	0,014	-0,293	0,024	0,784	-0,093	0,028
CT113	0,140	0,015	-0,911	0,026	0,257	-0,117	0,030
CT114	0,573	0,014	-0,580	0,025	0,539	0,034	0,029
CT115	0,151	0,015	-0,879	0,026	0,285	-0,134	0,030
CT116	1,116	0,013	0,254	0,024	1,250	-0,134	0,027

Tabla 4.18 Análisis de FDI de las medidas, prueba de Matemática de 2.º grado de secundaria, usadas en la equiparación de 2022 con 2023

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA001	1,133	0,013	0,080	0,014	1,132	0,001	0,019
MA002	2,199	0,011	1,224	0,015	2,242	-0,043	0,019
MA003	0,126	0,011	-0,539	0,014	0,532	-0,406	0,018
MA004	0,735	0,010	-0,338	0,011	0,727	0,008	0,015
MA005	0,792	0,013	-0,387	0,011	0,679	0,113	0,017
MA008	2,013	0,014	0,625	0,014	1,661	0,352	0,020
MA009	1,015	0,010	-0,058	0,014	0,998	0,017	0,017
MA011	0,971	0,008	-0,308	0,014	0,756	0,215	0,016
MA012	-0,908	0,018	-1,995	0,019	-0,881	-0,027	0,026

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA013	1,438	0,013	0,253	0,014	1,300	0,138	0,019
MA014	1,064	0,013	0,095	0,014	1,147	-0,083	0,019
MA015	2,072	0,014	1,385	0,011	2,398	-0,326	0,018
MA017	1,902	0,013	0,691	0,014	1,725	0,177	0,019
MA018	0,718	0,007	-0,121	0,014	0,937	-0,219	0,016
MA020	2,412	0,011	1,262	0,015	2,278	0,134	0,019
MA023	2,464	0,009	1,518	0,012	2,527	-0,063	0,015
MA024	2,009	0,011	0,841	0,014	1,870	0,139	0,018
MA030	0,837	0,013	-0,493	0,011	0,576	0,261	0,017
MA032	-0,266	0,015	-1,534	0,017	-0,433	0,167	0,023
MA035	0,775	0,013	-0,215	0,014	0,846	-0,071	0,019
MA036	1,160	0,008	0,089	0,014	1,141	0,019	0,016
MA038	1,655	0,010	0,943	0,014	1,969	-0,314	0,017
MA041	1,467	0,013	0,515	0,011	1,554	-0,087	0,017
MA044	1,766	0,012	0,884	0,014	1,912	-0,146	0,018
MA056	1,117	0,012	0,177	0,014	1,226	-0,109	0,018
MA057	1,605	0,013	0,482	0,014	1,522	0,083	0,019
MA076	0,803	0,013	-0,364	0,011	0,701	0,102	0,017
MA077	-0,185	0,008	-1,351	0,017	-0,256	0,071	0,019
MA078	0,360	0,014	-0,759	0,015	0,318	0,042	0,021
MA079	-0,313	0,016	-1,294	0,013	-0,201	-0,112	0,021
MA081	-0,101	0,012	-1,049	0,015	0,037	-0,138	0,019
MA082	1,149	0,013	-0,030	0,014	1,025	0,124	0,019
MA083	0,286	0,014	-0,666	0,014	0,408	-0,122	0,020
MA084	1,209	0,013	0,050	0,011	1,103	0,106	0,017
MA089	-0,186	0,015	-1,386	0,013	-0,290	0,104	0,020
MA092	1,748	0,013	0,714	0,014	1,747	0,001	0,019
MA097	1,327	0,013	0,213	0,014	1,261	0,066	0,019
MA099	0,806	0,013	-0,168	0,014	0,891	-0,085	0,019
MA100	-0,345	0,012	-1,258	0,016	-0,166	-0,179	0,020
MA102	0,903	0,010	-0,331	0,014	0,733	0,170	0,017
MA103	0,547	0,014	-0,372	0,014	0,694	-0,147	0,020
MA105	1,071	0,013	0,009	0,014	1,063	0,008	0,019
MA107	-0,715	0,017	-2,179	0,017	-1,059	0,344	0,024
MA108	0,525	0,014	-0,424	0,014	0,643	-0,118	0,020

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA111	1,449	0,013	0,310	0,011	1,355	0,094	0,017
MA113	1,466	0,013	0,423	0,014	1,465	0,001	0,019
MA114	1,919	0,013	0,751	0,014	1,783	0,136	0,019
MA115	0,644	0,013	-0,375	0,014	0,691	-0,047	0,019
MA116	1,161	0,013	0,264	0,011	1,310	-0,149	0,017
MA117	2,093	0,011	0,920	0,014	1,947	0,146	0,018
MA118	0,583	0,014	-0,174	0,014	0,886	-0,303	0,020
MA120	1,690	0,013	0,705	0,011	1,738	-0,048	0,017

Tabla 4.19 Análisis de FDI de las medidas, prueba de Ciencias Sociales de 2.º grado de secundaria, usadas en la equiparación de 2018 con 2023

Ítem	2018		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CS001	-2,103	0,008	-1,792	0,022	-1,843	-0,260	0,023
CS002	-0,821	0,006	-0,942	0,017	-0,951	0,130	0,018
CS006	1,420	0,006	1,046	0,016	1,135	0,285	0,017
CS008	0,118	0,005	0,043	0,006	0,083	0,035	0,008
CS009	-1,897	0,007	-1,555	0,020	-1,594	-0,303	0,021
CS011	0,689	0,005	0,664	0,016	0,734	-0,045	0,017
CS013	0,142	0,005	0,444	0,016	0,503	-0,361	0,017
CS014	-0,685	0,006	-0,765	0,017	-0,765	0,081	0,018
CS015	-0,275	0,005	-0,547	0,016	-0,536	0,261	0,017
CS034	-0,195	0,005	-0,468	0,016	-0,454	0,259	0,017
CS035	-0,533	0,005	-0,716	0,007	-0,714	0,180	0,009
CS036	-1,061	0,006	-1,097	0,018	-1,113	0,053	0,019
CS038	-0,966	0,006	-1,165	0,018	-1,185	0,219	0,019
CS039	-0,180	0,005	-0,357	0,007	-0,337	0,157	0,009
CS043	0,346	0,005	0,481	0,015	0,542	-0,197	0,016
CS044	-0,571	0,005	-0,636	0,016	-0,630	0,059	0,017
CS045	0,591	0,005	0,372	0,016	0,428	0,164	0,017
CS046	-1,365	0,006	-1,321	0,019	-1,349	-0,017	0,020
CS047	-1,048	0,006	-0,892	0,017	-0,898	-0,150	0,018
CS079	0,368	0,005	0,111	0,016	0,154	0,214	0,017
CS081	1,221	0,006	1,189	0,016	1,285	-0,064	0,017

Ítem	2018		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CS082	0,646	0,005	0,727	0,016	0,800	-0,154	0,017
CS085	0,600	0,005	0,637	0,016	0,706	-0,106	0,017
CS089	0,151	0,005	0,101	0,016	0,143	0,007	0,017
CS090	-0,390	0,005	-0,334	0,016	-0,313	-0,077	0,017
CS092	0,138	0,005	0,378	0,016	0,434	-0,296	0,017
CS093	0,849	0,005	0,690	0,016	0,761	0,087	0,017
CS094	-0,160	0,005	-0,482	0,016	-0,468	0,308	0,017
CS095	-0,368	0,005	-0,318	0,016	-0,296	-0,071	0,017
CS097	0,869	0,005	1,068	0,016	1,158	-0,289	0,017
CS099	0,156	0,005	0,216	0,006	0,264	-0,109	0,008

Tabla 4.20 Constantes de equiparación de las pruebas de la ENLA muestral 2023

Prueba	a	b
Lectura, 2.º grado de primaria	0,951	1,729
Matemática, 2.º grado de primaria	0,994	0,267
Lectura, 4.º grado de primaria	1,067	-0,414
Matemática, 4.º grado de primaria	1,084	-0,839
Lectura, 2.º grado de secundaria	0,852	1,033
Matemática, 2.º grado de secundaria	0,970	1,054
Ciencia Sociales, 2.º grado de secundaria	1,049	0,038

4.5 Estrategias de análisis psicométrico en educación intercultural bilingüe (EIB)

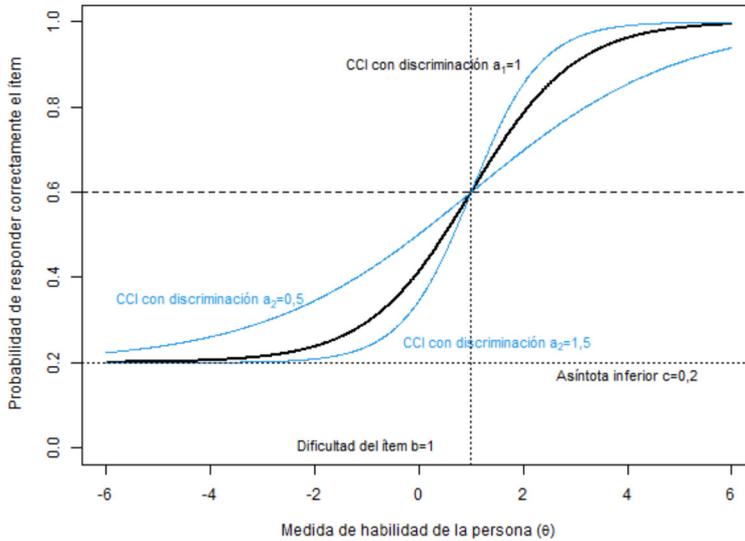
4.5.1 Modelo TRI logístico de tres parámetros

Para la población en IE EIB, se aplicó el modelo logístico que considera tres parámetros (Birnbbaum, en Lord y Novick, 1968). Este modelo asume que la probabilidad de responder correctamente un ítem se define, además de la diferencia entre la habilidad latente y la dificultad del ítem, por una medida de discriminación del ítem y una medida asíntota inferior denominada seudoadivinación. De esta forma, la curva característica del ítem (probabilidad de responder correctamente un ítem) está dada por la siguiente función matemática:

$$P_{ni} = P(X_{ni} = 1 | \theta_n, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{a_i(\theta_n - b_i)}}{1 + e^{a_i(\theta_n - b_i)}}$$

Donde θ_n representa el rasgo latente de la persona n , b_i simboliza el parámetro de dificultad del ítem i , a_i simboliza el parámetro de discriminación del ítem i y c_i simboliza el parámetro de pseudoaleatorización.

Figura 4.5 Curva características de ítem en el modelo TRI logístico de tres parámetros



La Figura 4.5 presenta tres curvas características de un ítem, las cuales consideran un mismo valor de dificultad $b=1$ y pseudoaleatorización $c=0.2$ con distintos valores del parámetro de discriminación a_i ; se observa que si $\theta > b_i$, entonces la probabilidad de responder correctamente el ítem (P_{ni}) se acerca más rápido a 1 a medida que a_i posee un valor mayor. Si $\theta < b_i$, la probabilidad de responder correctamente al ítem se acerca más rápido a la asíntota inferior a medida que a_i posee un valor más alto.

De acuerdo con De Ayala (2022), el motivo para tener en cuenta el parámetro de pseudoaleatorización es reflejar que algunos individuos con localizaciones infinitamente bajas puedan obtener una respuesta correcta. Estas respuestas son una manifestación de la interacción entre la persona y las características del ítem, lo que genera discusiones en torno a la naturaleza de este tercer parámetro. Además, las características personales incluyen no solo la habilidad evaluada, sino también sus tendencias a realizar pruebas y a “asumir riesgos”. Estos dos últimos factores son variables latentes tangenciales de la persona. Por lo tanto, aunque c se considera un parámetro del ítem, puede reflejar más una característica de la persona (es decir, un parámetro de la persona) que una característica del ítem o, al menos, una interacción entre las características de la persona y del ítem.

Es debido a lo anterior que este tercer parámetro recibe el nombre de seudoadivinación, ya que los aciertos en el ítem no se deben únicamente a la adivinanza, sino que pueden reflejar otros comportamientos o procesos de interacción entre el ítem y la persona. Por otro lado, a pesar de la vigente discusión en torno a la naturaleza del tercer parámetro (De Ayala, 2022), se priorizó su utilidad para informar sobre la probabilidad de responder correctamente a un ítem, incluso cuando los niveles de habilidad de quienes se enfrentan a él son bajos.

En términos matemáticos, la asíntota inferior describe el valor al que se aproxima la función a medida que θ se aproxima a infinito negativo. Por tanto, el parámetro c_i tiene un rango teórico de 0 a 1. Una particularidad de este modelo es que el valor de c_i no varía en función del nivel de habilidad θ ; es decir que, independientemente de la ubicación de una persona, su propensión a “adivinar” es constante en todo el continuo (De Ayala, 2022).

Considerar un tercer parámetro en el modelo conlleva otros efectos: los c_i no nulos disminuyen la estimación de la ubicación de una persona y reducen la cantidad de información del ítem. Por lo tanto, aunque se esté modelando los c_i no nulos, es deseable que los c_i sean cercanos a cero (De Ayala, 2022).

4.5.2 Análisis de ítems

Para analizar el ajuste del modelo en la estimación de los parámetros, se calculó el índice $S-X^2$ para cada ítem, así como la función de información del ítem, y se revisó los valores estimados del parámetro de discriminación.

El índice de bondad de ajuste $S-X^2$ (Orlando y Thissen, 2000) divide las respuestas de los evaluados en k grupos ordenados según la cantidad de respuestas correctas de cada persona. En cada grupo, compara las proporciones observadas con las proporciones esperadas y evalúa la hipótesis nula de si ambas son similares, para lo cual calcula una probabilidad ($p.S-X^2$) a partir de los datos. Esta probabilidad permite verificar si la hipótesis nula se rechaza o no. A continuación, se presenta la fórmula matemática del índice $S-X^2$

$$S - X_i^2 = \sum_{k=1}^{n-1} N_k \frac{(O_{ik} - E_{ik})^2}{E_{ik}(1 - E_{ik})}$$

Donde O_{ik} son las proporciones observadas para el ítem i en el grupo k obtenido según las puntuaciones de respuestas correctas. E_{ik} son las proporciones esperadas para el ítem i en el grupo k . N_k es el tamaño del grupo k . n es el número de grupos.

Además de evaluarse el ajuste de los ítems, se revisó la estimación de los valores del parámetro de discriminación de los ítems, ya que estos podrían no diferenciar adecuadamente entre examinados con diferentes niveles de habilidad. Baker y Kim

(2017) sugieren una clasificación del parámetro de discriminación indicando que un nivel moderado se establece a partir de 0,65.

4.5.3 Evidencias de validez vinculadas a la estructura interna de las pruebas

En relación con la evidencia de validez basada en la estructura interna, se evalúa el supuesto de unidimensionalidad considerando que, aunque no existe un instrumento que indique unidimensionalidad en su totalidad, se espera que un instrumento de medida logre representar las medidas en un solo factor dominante (Embretson y Reise, 2000). En ese sentido, se realiza un análisis de componentes principales de los residuos. Como se indicó con anterioridad, este análisis de los residuos sirve para detectar cualquier varianza que quede en los datos luego de extraer la dimensión principal (Bond et al., 2021). Este último análisis toma en cuenta los residuos estandarizados luego de ajustar al modelo de tres parámetros; estos residuos son calculados con la siguiente fórmula, la misma que está en línea con lo especificado en el reporte técnico de PISA (OCDE, 2022):

$$r(X_{ni}) = (X_{ni} - P_{ni}) / \sqrt{(P_{ni}(1 - P_{ni}))}$$

Donde $P_{ni} = P(X_{ni} = 1 | \hat{\theta}_n, a_i, b_i, c_i)$ es la probabilidad de acierto obtenida a partir del modelo logístico de tres parámetros (ver sección 4.5.1).

4.5.4 Confiabilidad y consistencia de la clasificación

Debido a la incorporación de nuevos parámetros al modelo psicométrico, no se puede utilizar el mismo coeficiente de separación de personas empleado bajo el modelo Rasch. En el caso de las pruebas que utilizan el modelo logístico de tres parámetros, se emplearon dos métodos para evaluar la fiabilidad de las puntuaciones: el coeficiente de fiabilidad empírica y la función de información de la prueba. La justificación para el uso de más de un método se encuentra en el estándar 2.14 de los Estándares para Pruebas Educativas y Psicológicas (American Educational Research Association [AERA] et al., 2014), que recomienda reportar errores estándar condicionales en múltiples niveles de puntuación. Por lo tanto, además de disponer de un estadístico como medida representativa de la fiabilidad, es igualmente conveniente conocer los errores estándar, ya que proporcionan un mayor detalle sobre la precisión de las puntuaciones de la prueba a lo largo de la escala.

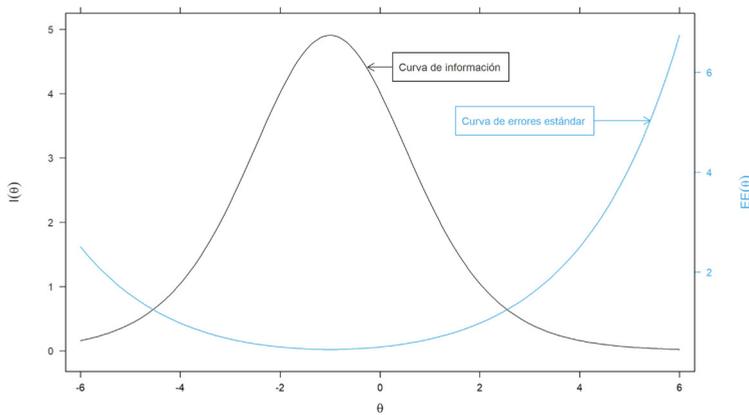
El coeficiente de fiabilidad empírica (Bock y Zimowski, en du Toit, 2003, pp. 33-34) deriva su naturaleza e interpretación de la definición clásica de la fiabilidad en la TCT: la relación entre la varianza de la puntuación verdadera y la varianza de la puntuación observada, siendo esta última la suma de la varianza de las puntuaciones verdaderas

y la varianza del error. En el contexto de la TRI, las puntuaciones verdaderas son los valores θ no observables que se estiman con un error estándar especificado a partir de los patrones de respuesta. El cálculo del coeficiente lleva la siguiente fórmula:

$$fiabilidad = \frac{VAR(\theta)}{VAR(\theta) + VAR(error)}$$

Donde $VAR(\theta)$ es la varianza de los valores θ verdaderos y $VAR(error)$ es la media aritmética de los errores estándar luego de haber sido elevados al cuadrado.

Figura 4.6 Curva de información de la prueba y curva de errores estándar.



En la TRI, la función de información de la prueba se utiliza para calcular el error estándar de medición y, por ende, también puede emplearse para generar coeficientes con interpretaciones similares a los coeficientes de confiabilidad (AERA et al., 2014; DeMars, 2010). La información de la prueba, $I(\theta)$, es una función del rasgo evaluado y de los ítems que lo componen, proporcionando una medida de la precisión en cada nivel de habilidad. Estas funciones de información de la prueba se representan gráficamente en una curva de información, que se muestra en color negro en la figura 4.6. En dicha figura, se observa que hay más información, principalmente entre los valores -2 y 0 de la escala, y menor información fuera de este rango.

Además, la información de la prueba refleja que cada uno de los ítems contribuye potencialmente a reducir la incertidumbre sobre la localización de una persona, independientemente de los demás ítems del instrumento. Gracias a esta independencia entre ítems, es posible sumar las contribuciones de cada uno para obtener la información (total) de la prueba (De Ayala, 2022). La contribución individual de cada ítem se conoce como información del ítem, $I_i(\theta)$.

De este modo, el cálculo de la función de información de la prueba queda definida como:

$$I(\theta) = \sum I_i(\theta)$$

Y el cálculo de la función de información de un ítem j bajo el modelo logístico de tres parámetros es:

$$I_i(\theta) = a_i^2 \left[\frac{(\theta - c_i)^2}{(1 - c_i)^2} \right] \left[\frac{1 - p_i}{p_i} \right]$$

Donde p_i representa la probabilidad de acertar al ítem i en cierto nivel θ , a_i es el parámetro de discriminación del ítem i , c_i es el parámetro de seudoadivinación del ítem i .

Como se observa, la información de un ítem depende de sus parámetros. La información del ítem es más alta cuando el parámetro a es elevado y más baja cuando el parámetro a es menor. Además, en el caso de ítems dicotómicos evaluados bajo el modelo logístico de tres parámetros, la información no alcanza su valor máximo en un nivel de habilidad correspondiente al parámetro de dificultad ($\theta = b$). Esto se debe a que un valor distinto de cero para el parámetro c disminuye la información, especialmente para los examinados con baja capacidad (Baker y Kim, 2017). Esto conlleva a un desplazamiento del punto máximo de información ligeramente por encima del parámetro de dificultad (DeMars, 2010).

Por su parte, el error estándar de medición (o error estándar de θ) para las puntuaciones es el inverso de la raíz cuadrada de la función de información de la prueba:

$$EE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

De este modo, a mayor información, menor es el error estándar y, por ende, mayor es la fiabilidad en dicho nivel θ . El conjunto de errores estándar también se representa mediante una curva denominada curva de errores estándar, y esta relación inversa se puede observar también en la figura 4.6. Por ejemplo, mientras la curva de información decrece aproximadamente a partir del valor cero de la escala, la curva de errores estándar, mostrada en color celeste, crece. La información se incrementa con cada ítem adicional, pero el error estándar de medición no disminuye proporcionalmente ya que la función de raíz cuadrada es una transformación no lineal (DeMars, 2010).

4.6 Resultados del análisis psicométrico en Lectura castellano como segunda lengua EIB

4.6.1 Calibración de los ítems

En la tabla 4.21 se presenta los parámetros estimados, los errores de estimación asociados, el valor de p para el índice $S-X^2$ para cada uno de los ítems de la prueba de Lectura en castellano como segunda lengua aplicada a los estudiantes de 4.º grado de primaria que asisten a instituciones educativas EIB.

Tabla 4.21 Parámetros estimados y ajuste de los ítems al modelo logístico de tres parámetros, Castellano como segunda lengua de 4.º grado de primaria

Ítem	Bloque	Posición	b	Error (b)	a	Error (a)	c	Error (c)	$p, S-X^2$	Nivel
L2001	B07	2	-0,782	0,021	2,948	0,076	0,246	0,013	0,365	2
L2002	B18	4	-0,370	0,017	3,619	0,099	0,355	0,009	0,765	3
L2003	B18	5	-0,319	0,011	4,426	0,101	0,152	0,006	0,803	3
L2004	B18	6	0,039	0,011	3,961	0,091	0,164	0,005	0,480	4
L2005	B04	1	-1,802	0,090	1,748	0,059	0,077	0,068	0,803	1
L2006	B12	1	0,330	0,012	3,555	0,099	0,235	0,005	0,736	4
L2007	B12	2	0,313	0,011	3,999	0,106	0,179	0,005	0,846	4
L2008	B12	3	0,215	0,021	2,101	0,060	0,299	0,008	0,717	4
L2009	B20	1	1,144	0,015	3,734	0,142	0,334	0,004	0,439	5
L2010	B20	3	1,394	0,026	1,606	0,064	0,189	0,007	0,439	5
L2011	B20	4	1,815	0,025	4,519	0,258	0,289	0,003	0,508	5
L2012	B01	5	-1,269	0,020	3,005	0,069	0,021	0,014	0,434	2
L2013	B08	1	-0,644	0,013	4,013	0,094	0,145	0,008	0,976	3
L2014	B10	1	-0,264	0,011	4,361	0,112	0,184	0,006	0,363	3
L2015	B10	3	-0,213	0,016	2,848	0,070	0,215	0,008	0,508	3
L2016	B10	4	0,980	0,013	3,768	0,145	0,323	0,004	0,750	5
L2017	B17	6	-1,226	0,024	2,234	0,046	0,002	0,015	0,717	2
L2018	B21	2	0,032	0,013	4,399	0,127	0,349	0,006	0,582	4
L2019	B21	3	0,840	0,014	3,560	0,112	0,289	0,005	0,506	5
L2020	B07	5	0,119	0,021	2,052	0,055	0,239	0,009	0,765	4
L2021	B07	6	0,000	0,020	2,236	0,058	0,255	0,009	0,738	4
L2022	B21	1	-0,151	0,013	3,535	0,083	0,204	0,007	0,736	3
L2023	B21	4	0,065	0,012	3,458	0,075	0,136	0,005	0,465	4
L2024	B14	1	0,086	0,019	2,683	0,079	0,367	0,008	0,376	4
L2025	B14	4	0,004	0,020	1,960	0,048	0,150	0,009	0,736	4
L2026	B14	3	0,814	0,014	2,841	0,085	0,208	0,005	0,423	4

Ítem	Bloque	Posición	<i>b</i>	Error (<i>b</i>)	<i>a</i>	Error (<i>a</i>)	<i>c</i>	Error (<i>c</i>)	$p, S-X^2$	Nivel
L2027	B14	5	1,382	0,025	1,547	0,070	0,213	0,007	0,859	5
L2028	B05	1	-0,787	0,015	4,050	0,106	0,217	0,010	0,803	2
L2029	B19	2	-0,397	0,012	4,867	0,122	0,196	0,007	0,968	3
L2030	B05	6	1,435	0,017	4,143	0,204	0,324	0,004	0,803	5
L2031	B23	2	0,826	0,013	2,902	0,077	0,169	0,005	0,846	4
L2032	B23	4	0,925	0,045	0,834	0,032	0,020	0,015	0,736	5
L2033	B23	5	0,049	0,012	3,510	0,081	0,183	0,006	0,423	4
L2034	B23	6	1,381	0,026	1,866	0,078	0,301	0,006	0,717	5
L2035	B15	3	0,794	0,026	1,636	0,062	0,294	0,009	0,836	4
L2036	B15	6	0,934	0,017	2,392	0,081	0,273	0,006	0,363	5
L2037	B15	1	0,395	0,013	3,461	0,095	0,237	0,005	0,615	4
L2038	B15	4	0,141	0,023	1,783	0,047	0,167	0,010	0,683	4
L2039	B15	5	1,888	0,036	2,388	0,156	0,359	0,005	0,439	5
L2040	B15	2	1,147	0,014	3,498	0,143	0,275	0,004	0,711	5
L2041	B18	1	-0,432	0,011	5,984	0,180	0,267	0,007	0,942	3
L2042	B09	1	-0,012	0,018	2,939	0,087	0,397	0,008	0,439	4
L2043	B09	4	0,560	0,014	3,384	0,116	0,380	0,005	0,968	4
L2044	B09	5	-0,044	0,013	4,190	0,124	0,342	0,007	0,673	3
L2045	B13	1	0,702	0,018	2,378	0,075	0,294	0,006	0,942	4
L2046	B13	2	0,087	0,014	3,583	0,096	0,275	0,006	0,618	4
L2047	B13	3	1,278	0,025	1,800	0,074	0,287	0,007	0,599	5
L2048	B13	4	0,694	0,018	2,158	0,064	0,233	0,006	0,717	4
L2049	B18	2	-0,476	0,012	5,161	0,134	0,191	0,007	0,480	3
L2050	B22	1	0,476	0,015	3,145	0,090	0,315	0,006	0,846	4
L2051	B22	2	0,181	0,024	1,701	0,043	0,140	0,010	0,727	4
L2052	B22	3	-	-	-	-	-	-	-	-
L2053	B22	4	1,076	0,018	2,585	0,091	0,309	0,005	0,717	5
L2054	B22	5	0,913	0,024	1,534	0,051	0,171	0,008	0,968	5
L2055	B22	6	1,529	0,022	2,855	0,121	0,298	0,004	0,465	5
L2056	B08	2	-0,520	0,015	4,368	0,128	0,353	0,009	0,962	3
L2057	B16	3	-1,494	0,034	2,436	0,063	0,010	0,026	0,721	1
L2058	B17	2	-1,097	0,022	2,281	0,046	0,003	0,013	0,514	2
L2059	B17	3	-1,660	0,044	2,283	0,063	0,004	0,036	0,363	1
L2060	B02	1	-1,096	0,015	3,199	0,064	0,005	0,008	0,439	2
L2061	B16	2	-1,646	0,066	1,792	0,053	0,008	0,048	0,802	1

Ítem	Bloque	Posición	<i>b</i>	Error (<i>b</i>)	<i>a</i>	Error (<i>a</i>)	<i>c</i>	Error (<i>c</i>)	<i>p</i> , S-X ²	Nivel
L2062	B04	2	-1,271	0,022	2,286	0,045	0,002	0,012	0,615	2
L2063	B02	2	-1,165	0,021	3,614	0,099	0,194	0,016	0,846	2
L2064	B02	3	-1,045	0,014	4,777	0,126	0,115	0,010	0,942	2
L2065	B06	2	-0,602	0,014	4,807	0,144	0,316	0,009	0,721	3
L2066	B08	4	-0,271	0,017	3,125	0,081	0,331	0,009	0,433	3
L2067	B08	5	-0,329	0,012	4,585	0,114	0,199	0,007	0,674	3
L2068	B05	2	-0,603	0,013	4,333	0,110	0,189	0,008	0,439	3
L2069	B16	1	-1,250	0,029	2,557	0,064	0,122	0,021	0,615	2
L2070	B17	1	-1,878	0,071	1,847	0,055	0,005	0,056	0,423	1
L2071	B04	5	-1,172	0,015	3,055	0,060	0,002	0,007	0,363	2
L2072	B02	7	-0,799	0,019	3,544	0,097	0,272	0,012	0,363	2
L2073	B01	8	-0,834	0,016	2,865	0,059	0,053	0,010	0,567	2
L2074	B01	1	-1,303	0,022	2,441	0,051	0,002	0,014	0,615	2
L2075	B04	3	-1,155	0,015	2,918	0,055	0,001	0,006	0,615	2
L2076	B04	8	-0,783	0,022	2,745	0,069	0,238	0,013	0,363	2
L2077	B07	1	-0,665	0,021	3,156	0,089	0,330	0,012	0,717	3
L2078	B19	1	-0,427	0,013	4,389	0,110	0,227	0,008	0,830	3
L2079	B19	4	-0,156	0,010	5,662	0,142	0,153	0,005	0,674	3
L2080	B19	5	0,105	0,015	3,439	0,094	0,333	0,007	0,717	4
L2081	B19	6	-0,369	0,016	2,827	0,060	0,132	0,008	0,803	3
L2082	B06	1	-0,692	0,017	4,233	0,128	0,344	0,010	0,846	3
L2083	B12	4	0,223	0,012	3,498	0,090	0,194	0,005	0,508	4
L2084	B12	5	0,281	0,019	1,753	0,045	0,105	0,008	0,433	4
L2085	B13	5	0,723	0,024	1,452	0,045	0,127	0,009	0,717	4
L2086	B13	6	0,750	0,026	1,546	0,053	0,226	0,009	0,968	4
L2087	B01	3	-1,582	0,029	2,384	0,056	0,002	0,021	0,674	1
L2088	B01	2	-1,758	0,055	2,246	0,070	0,012	0,048	0,618	1
L2089	B01	4	-1,301	0,020	3,044	0,070	0,009	0,013	0,439	2
L2090	B01	6	-1,084	0,025	2,165	0,046	0,005	0,015	0,439	2
L2091	B01	7	-0,776	0,017	2,975	0,065	0,111	0,011	0,717	3
L2092	B05	3	-0,267	0,012	4,322	0,104	0,188	0,007	0,439	3
L2093	B05	4	0,079	0,010	5,934	0,170	0,190	0,005	0,674	4
L2094	B05	5	-0,175	0,011	5,076	0,135	0,224	0,006	0,514	3
L2095	B09	2	-0,113	0,013	3,729	0,094	0,250	0,007	0,742	3
L2096	B09	3	0,060	0,013	4,126	0,121	0,332	0,006	0,859	4

Ítem	Bloque	Posición	<i>b</i>	Error (<i>b</i>)	<i>a</i>	Error (<i>a</i>)	<i>c</i>	Error (<i>c</i>)	<i>p,S-X²</i>	Nivel
L2097	B11	1	1,236	0,014	3,989	0,162	0,266	0,004	0,465	5
L2098	B11	2	0,820	0,015	2,624	0,083	0,262	0,005	0,968	4
L2099	B11	3	0,960	0,015	2,935	0,105	0,297	0,005	0,502	5
L2100	B11	4	0,565	0,012	4,554	0,172	0,376	0,005	0,803	4
L2101	B11	5	0,538	0,015	3,113	0,100	0,316	0,006	0,363	4
L2102	B02	4	-1,018	0,015	5,108	0,148	0,172	0,011	0,674	2
L2103	B02	5	-0,809	0,018	3,242	0,078	0,182	0,012	0,859	2
L2104	B02	6	-1,058	0,016	3,479	0,076	0,043	0,010	0,665	2
L2105	B02	8	-0,766	0,022	3,184	0,088	0,324	0,013	0,962	3
L2106	B06	3	0,238	0,012	4,095	0,114	0,292	0,006	0,363	4
L2107	B06	4	-0,537	0,017	3,367	0,086	0,258	0,010	0,674	3
L2108	B06	5	0,383	0,033	1,534	0,054	0,307	0,011	0,363	4
L2109	B06	6	-0,049	0,015	3,596	0,100	0,339	0,007	0,803	3
L2110	B10	2	1,328	0,023	1,592	0,068	0,191	0,007	0,555	5
L2111	B10	5	0,295	0,046	0,923	0,033	0,006	0,017	0,982	4
L2112	B14	2	0,181	0,021	1,842	0,048	0,160	0,009	0,803	4
L2113	B14	6	0,523	0,014	3,100	0,091	0,309	0,006	0,717	4
L2114	B03	1	-1,314	0,023	3,159	0,080	0,078	0,017	0,846	2
L2115	B03	2	-1,535	0,027	2,956	0,080	0,030	0,023	0,711	1
L2116	B03	3	-1,256	0,028	2,577	0,064	0,092	0,020	0,615	2
L2117	B03	4	-1,400	0,025	2,953	0,075	0,054	0,019	0,803	2
L2118	B03	5	-1,320	0,023	3,633	0,106	0,182	0,019	0,439	2
L2119	B03	6	-1,380	0,023	3,028	0,075	0,022	0,017	0,615	2
L2120	B03	7	-1,127	0,037	2,290	0,065	0,255	0,022	0,968	2
L2121	B03	8	-0,873	0,063	1,162	0,035	0,007	0,029	0,736	2
L2122	B07	3	-0,140	0,016	3,015	0,076	0,258	0,008	0,433	3
L2123	B07	4	0,410	0,024	1,636	0,047	0,193	0,010	0,514	4
L2124	B04	4	-1,148	0,016	3,092	0,063	0,005	0,008	0,423	2
L2125	B04	6	-1,091	0,017	2,537	0,048	0,002	0,009	0,736	2
L2126	B04	7	-0,885	0,017	3,219	0,072	0,092	0,011	0,717	2
L2127	B08	3	-0,088	0,012	5,510	0,175	0,349	0,006	0,433	3
L2128	B08	6	-0,083	0,041	1,235	0,037	0,097	0,017	0,711	3
L2129	B20	2	0,370	0,047	1,257	0,048	0,273	0,015	0,711	4
L2130	B20	5	-	-	-	-	-	-	-	-
L2131	B21	5	0,512	0,025	1,391	0,038	0,077	0,010	0,711	4

Ítem	Bloque	Posición	<i>b</i>	Error (<i>b</i>)	<i>a</i>	Error (<i>a</i>)	<i>c</i>	Error (<i>c</i>)	<i>p</i> , <i>S</i> - <i>X</i> ²	Nivel
L2132	B16	4	-1,049	0,019	2,425	0,048	0,003	0,011	0,433	2
L2133	B16	5	-1,073	0,016	2,903	0,058	0,005	0,009	0,434	2
L2134	B16	6	-1,006	0,017	2,821	0,057	0,011	0,010	0,765	2
L2135	B16	7	-1,326	0,031	2,282	0,055	0,017	0,023	0,859	2
L2136	B16	8	-0,966	0,016	2,754	0,052	0,005	0,008	0,717	2
L2137	B18	3	-0,039	0,015	2,784	0,062	0,167	0,007	0,711	4
L2138	B17	4	-0,918	0,014	3,003	0,056	0,005	0,007	0,861	2
L2139	B17	5	-0,951	0,015	3,921	0,092	0,097	0,010	0,433	2
L2140	B17	7	-0,829	0,014	3,652	0,077	0,061	0,008	0,846	2
L2141	B17	8	-1,138	0,038	1,768	0,042	0,007	0,023	0,859	2
L2142	B19	3	-0,283	0,021	1,549	0,030	0,004	0,009	0,803	3
L2143	B23	1	0,242	0,012	3,738	0,096	0,243	0,006	0,363	4
L2144	B23	3	0,081	0,034	1,325	0,037	0,087	0,014	0,665	4

Nota: *a* = Discriminación; *b* = Dificultad; *c* = Seudoadivinación

Los ítems L2052 y L2130 de la prueba de Lectura en castellano como segunda lengua de 4.º grado de primaria EIB no presentaron un buen ajuste al modelo logístico de tres parámetros. Tales ítems no fueron incluidos en la estimación de las medidas de habilidad de los estudiantes evaluados.

4.6.2 Evidencias de validez vinculadas a la estructura interna

Para la prueba de Lectura en castellano como segunda lengua de 4.º grado de primaria EIB, se calculó los residuos estandarizados luego de ajustar el modelo, y se realizó un análisis de componentes principales para verificar si los residuos presentan alguna asociación de ítems. Los resultados se presentan en la tabla 4.22.

Tabla 4.22 Análisis de componentes principales de las medidas derivadas de las pruebas de 4.º grado de primaria de EIB

Prueba	Varianza de la dimensión principal (%)	Varianza modelada (%)	Primer autovalor	Primer autovalor (%)
Lectura, 4.º grado de primaria	95,5	99,7	1,481	1,043

Siguiendo de forma análoga el razonamiento de Linacre (2024), el primer autovalor contiene menos del 5 % de varianza y su valor es menor que 3. Entonces, podemos afirmar que no existen evidencias serias que atenten contra el supuesto de unidimensionalidad. Además, podemos observar que la cantidad de varianza empírica es similar a la cantidad de varianza predicha por el modelo, lo cual nos

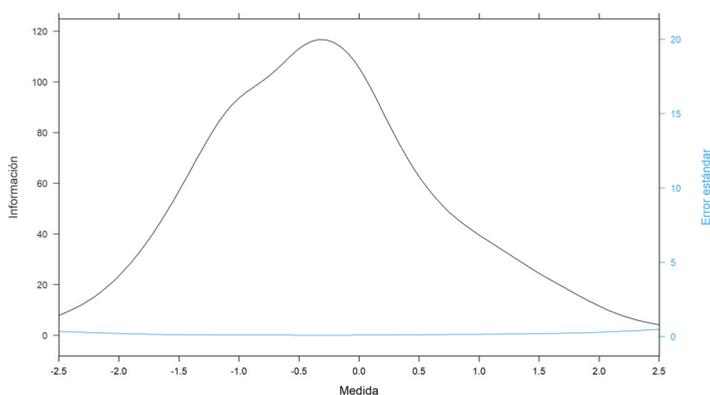
permite afirmar que no hay evidencias contundentes en contra del supuesto de unidimensionalidad del conjunto de ítems incluidos en la prueba de Lectura en castellano como segunda lengua.

4.6.3 Confiabilidad y consistencia de la clasificación

Para expresar los resultados de estos análisis, primero se utilizó el índice de fiabilidad empírica (ρ) como medida representativa de la confiabilidad de las puntuaciones. La prueba de Lectura en castellano como segunda lengua presentó un valor del índice de fiabilidad de $\rho=0,92$. Considerando este valor, se puede apreciar que la varianza de error de la prueba de Lectura es de 8 %. Por lo tanto, es posible afirmar que las medidas derivadas de estas pruebas poseen adecuadas evidencias de confiabilidad.

Como segundo criterio de evaluación de la confiabilidad de las puntuaciones, se muestran en la figura 4.7 la curva de información de la prueba y la curva de errores estándar.

Figura 4.7 *Curva de información de la prueba y curva de errores estándar de la prueba de Lectura en castellano como segunda lengua para 4.º grado de primaria EIB*



A partir de lo mostrado en la figura 4.7, puede concluirse que hay más información (es decir, la medida es más precisa) en los valores centrales de la escala, mientras que la información es menor en las puntuaciones extremas. Dado que la prueba posee un alto número de ítems, el nivel general de la función de información de la prueba es elevado, lo que proporciona estimaciones confiables en un amplio rango de medidas. Los errores estándar se mantienen por debajo de una unidad, incrementándose en los valores extremos de la escala.

Adicionalmente, se han agregado indicadores de precisión y consistencia de la clasificación de las medidas de habilidad de los estudiantes de EIB.

Tabla 4.23 Indicadores del análisis de precisión y consistencia de la clasificación usando las medidas derivadas de la prueba de Lectura en castellano como segunda lengua para 4.º grado de primaria EIB

	Lectura	
	Precisión	Consistencia
Nacional	0,77	0,67
Hombre	0,77	0,67
Mujer	0,77	0,67
Polidocente Completo	0,77	0,68
Unidocente / Multigrado	0,76	0,67
Lenguas andinas	0,77	0,68
Lenguas amazónicas	0,75	0,65

4.7 Niveles de logro y preparación de resultados

El diseño de las evaluaciones permite estimar, a partir de su desempeño en las pruebas, lo que los estudiantes saben y hacen respecto de lo que deberían saber y hacer. De acuerdo con ello, la interpretación de los resultados de la ENLA está referida a criterios o estándares de logro (Muñiz, 1998). Para hacerlo, es necesario establecer puntos de corte, lo que constituye un procedimiento estándar utilizado para delimitar numéricamente dos o más niveles de desempeño de una competencia (Cizek y Bunch, 2007). Estos niveles de desempeño, conceptualmente, corresponden a un juicio compartido por una comunidad de expertos (docentes, evaluadores, tomadores de decisiones), la cual determina qué es lo que un estudiante mínimamente competente debe poder hacer para ser considerado parte de un determinado nivel en pruebas referidas a criterios (Cizek, 1993; Shepard, 1980).

Este procedimiento se realiza por única vez al inicio de un ciclo de evaluaciones, con la intención de sostener los mismos puntos de corte en las ediciones posteriores y así asegurar que los resultados sean comparables en el tiempo. Este año, ha sido utilizado para determinar los niveles de la prueba de Lectura en castellano como segunda lengua en 4.º grado de primaria EIB.

Para establecer puntos de corte, se utiliza el método Bookmark (Cizek y Bunch, 2007; Karantonis y Sireci, 2006; Lewis et al., 1996), el cual es consistente con el modelo de medición utilizado por la UMC en las distintas evaluaciones que conduce. Como consecuencia, el juicio sobre los cortes considera también los ítems de respuesta construida y con créditos parciales, dado que estos se incluyen en la misma métrica que los de opción múltiple. El método Bookmark, en términos generales, implica colocar marcas (tantas como cortes se hayan preestablecido) en un cuadernillo de ítems ordenado por dificultad (la cual es estimada mediante procedimientos Rasch o de teoría de respuesta al ítem). La pregunta típica que guía el establecimiento de

cortes es la siguiente: “¿Hasta qué ítem debe ser capaz de resolver un estudiante, como mínimo, para ser considerado parte del nivel?”. El procedimiento establece que los jueces, organizados en grupos pequeños, determinen los cortes para cada nivel de desempeño; de no llegarse a un acuerdo, se aplican procedimientos estadísticos para resolver la discrepancia.

En general, un taller se conduce en tres rondas (Cizek y Bunch, 2007). En la primera ronda, los participantes leen las descripciones de los niveles de logro elaboradas por el equipo de especialistas (para el caso, de la UMC), resuelven todos los ítems de la prueba y analizan las razones por las cuales un ítem era más difícil que el anterior. Esta ronda concluye con un primer establecimiento individual de cortes. En la segunda ronda, los participantes exponen, en los subgrupos, las razones que los motivaron a colocar sus cortes. Asimismo, reciben un reporte de discrepancias en el que se señala qué tan distintos son sus juicios respecto de los demás grupos. Esta ronda finaliza con un segundo establecimiento individual de cortes. En la tercera ronda, los participantes tienen acceso al impacto de los datos (es decir, a la distribución de personas en los distintos niveles de desempeño) si los resultados de la segunda ronda hubiesen sido los definitivos. Con esta información, los jueces emiten su tercer y último corte.

Como ha sido mencionado, el modelo Rasch estima de manera conjunta la dificultad de los ítems y las medidas de habilidad de las personas. Utilizando dicha información, es posible calcular la probabilidad de un estudiante con una habilidad específica de responder correctamente un ítem que posee una dificultad concreta (P_{ni}).

A partir de lo trabajado en la Evaluación Nacional de 2004 (Ministerio de Educación [Minedu], 2005), se establece que un estudiante está en un nivel de logro si tiene una probabilidad mayor o igual a 0,62 de responder correctamente el ítem que marca el corte entre dos niveles de logro consecutivos. En este sentido, es importante considerar que la probabilidad de 0,62 para responder correctamente un ítem implica una medida de habilidad, la cual es medio logit superior a la medida de dificultad del ítem en cuestión. Una vez hecha esta adición para cada estudiante evaluado, se transforman los valores a medida 500 (M500) y se redondean a dos decimales. Finalmente, se asigna a cada estudiante a un determinado nivel de logro según su M500.

En la tabla 4.25, se presentan los puntos de corte (en M500) de las evaluaciones ENLA muestral 2023.

Tabla 4.24 Medidas (en M500) que marcan los puntos de corte de los alumnos en la ENLA muestral 2023

Área	< Nivel 1 vs, Nivel 1	Nivel 1 vs, Nivel 2	Nivel 2 vs Nivel 3
Lectura, 2.º grado de primaria	458,39	583,66	-
Matemática, 2.º grado de primaria	512,22	639,21	-
Lectura, 4.º grado de primaria	356,92	444,72	522,03
Matemática, 4.º grado de primaria	351,90	422,21	526,46
Lectura, 2.º grado de secundaria	505,14	580,61	641,25
Matemática, 2.º grado de secundaria	519,67	595,96	649,38
Ciencia Sociales, 2.º grado de secundaria	425,93	500,05	607,28

Finalmente, es importante señalar que las medidas de habilidad de los estudiantes, así como los puntos de corte antes mencionados, fueron transformadas linealmente ($Y = a + bX$), de tal manera que, en la primera evaluación que se realizó en cada grado y área, la media aritmética fue 500 y la desviación estándar fue 100. Para ello, se usaron las constantes que aparecen en la tabla 4.25.

Tabla 4.25 Valores de a y b utilizados para la transformación lineal de las medidas de la ENLA muestral 2023

Área	a	b
Lectura, 2.º grado de primaria	423,614	63,585
Matemática, 2.º grado de primaria	468,321	72,440
Lectura, 4.º grado de primaria	474,896	73,414
Matemática, 4.º grado de primaria	463,157	73,781
Lectura, 2.º grado de secundaria	474,896	73,414
Matemática, 2.º grado de secundaria	463,157	73,781
Ciencia Sociales, 2.º grado de secundaria	476,687	100,063

En relación con los niveles de la ENLA 2023 para la prueba de Lectura en castellano como segunda lengua aplicada en estudiantes de 4.º grado de primaria en instituciones educativas EIB, se determinaron, en un taller de puntos de corte llevado a cabo en la UMC, los puntos de corte bajo la estimación de las medidas del modelo logístico de tres parámetros. En este sentido, las decisiones para la selección de los puntos de corte fueron tomadas a partir de la revisión y evaluación de los parámetros estimados de dificultad, discriminación y seudoadivinación para cada ítem. Tomando en cuenta lo establecido por la Evaluación Nacional de 2004 (Minedu, 2005), se establece que un estudiante se encuentra en un nivel de logro si tiene una probabilidad mayor o igual a 0,62 de responder correctamente el ítem que marca el corte entre dos niveles de logro consecutivos. En el modelo de tres parámetros, la probabilidad de

acertar un ítem en el punto de dificultad b_i se incrementa en comparación al modelo Rasch. Esto sucede según el valor estimado del parámetro de pseudoadivinación. La siguiente fórmula permite calcular la probabilidad mencionada.

$$P_i = c_i + (1 - c_i)/2$$

Donde c_i es el parámetro estimado de pseudoadivinación para el ítem i .

Asimismo, en el modelo logístico de tres parámetros, la medida de habilidad con una probabilidad de 0,62 de acertar el ítem u en el punto de su dificultad b_u esta dado por la siguiente fórmula:

$$\hat{\theta}_u = b_u + \frac{1}{a_u} \ln\left(\frac{0,62 - c_u}{1 - 0,62}\right)$$

Donde a_u , b_u y c_u son los valores estimados en el modelo para los parámetros del ítem u ; este ítem es definido como umbral para pasar de un nivel a otro según los jueces que participaron del taller de puntos de corte.

En la tabla 4.26, se presentan los puntos de corte (en M500) de la ENLA 2023 para la prueba de Lectura en castellano como segunda lengua aplicada en 4.º grado de primaria.

Tabla 4.26 Medidas (en M500) que marcan los puntos de corte de los alumnos en la ENLA 2023 para 4.º grado de primaria EIB

Área	< Nivel 1 vs. Nivel 1	Nivel 1 vs. Nivel 2	Nivel 2 vs Nivel 3	Nivel 3 vs Nivel 4
Lectura en castellano como segunda lengua, 4.º grado de primaria	370,00	432,30	502,40	584,00

4.8 Análisis psicométricos para habilidades socioemocionales

En relación con el análisis psicométrico de las escalas de habilidades socioemocionales (HSE), se aplicó el modelo de escala de valoración de Rasch (Andrich, 1978) en su versión multidimensional para las escalas de toma de decisiones y resolución de problemas.

Las medidas de habilidad de las personas fueron estimadas con la metodología de valores plausibles. Se emplearon cinco de ellos por cada persona evaluada, pues se ha demostrado la superioridad de esta metodología para reportar estimaciones de la población (Wu, 2005). La estimación de estos valores plausibles consideró una

regresión latente, usando como covariables el sexo de los estudiantes, el tipo de gestión de la IE y el área geográfica en la cual se ubica cada IE.

4.8.1 Calibración de los ítems

En las tablas 1 y 2 se presentan los resultados de la calibración de los ítems de cada una de las escalas de HSE. La columna denominada “Medida” representa la dificultad del ítem. Así, una menor medida implica que el ítem es más fácil de endosar (es decir, es más fácil que los estudiantes respondan marcando alguna de las categorías de respuesta más altas), mientras que una mayor medida implica que el ítem es más difícil de endosar (es decir, es más difícil que los estudiantes respondan marcando alguna de las categorías de respuesta más altas). Asimismo, la dificultad de cada ítem viene acompañada de un error que representa qué tan precisa ha sido la calibración de cada ítem. Así, mientras más pequeño es el error, más precisa ha sido la calibración de la dificultad del ítem (y viceversa). Como se puede observar en ambas tablas, los errores de calibración son muy pequeños, lo que indica que la calibración de la dificultad fue bastante precisa. Por último, en las tablas mencionadas se observa que el infit y el outfit de todos los ítems de las escalas de HSE utilizadas en la ENLA 2023 mostraron valores dentro del rango aceptable, por lo que se concluye que existe evidencia de un buen ajuste de los datos al modelo Rasch.

Tabla 4.27 Calibración de los ítems de las escalas de HSE aplicadas en 2.º grado de secundaria

Escala	Subescala	Ítem	Medida	Error	Infit	Outfit
Toma de decisiones	Otros estilos	HSETDD2S001	0,198	0,003	1,02	1,03
		HSETDD2S007	0,273	0,003	1,00	1,01
		HSETDD2S015	0,075	0,003	0,94	0,94
		HSETDD2S004	-0,128	0,003	0,89	0,90
		HSETDD2S016	0,596	0,003	0,90	0,89
		HSETDD2S014	0,756	0,003	1,12	1,10
		HSETDD2S017	0,703	0,003	1,17	1,17
		HSETDD2S010	0,753	0,003	1,18	1,21
	Estilo vigilante	HSETDD2S003	-0,653	0,003	0,99	1,00
		HSETDD2S011	-0,385	0,003	0,95	0,95
		HSETDD2S013	-0,739	0,003	0,84	0,83
		HSETDD2S012	-0,893	0,003	1,05	1,04

Escala	Subescala	Ítem	Medida	Error	Infit	Outfit
Resolución de problemas	Confianza	HSEREP2S009	-0,790	0,004	1,02	1,02
		HSEREP2S011	-0,854	0,004	0,88	0,88
		HSEREP2S012	-1,075	0,004	0,94	0,93
		HSEREP2S013	-0,870	0,004	0,89	0,87
		HSEREP2S014	-0,768	0,004	0,83	0,83
		HSEREP2S019	-0,525	0,004	0,97	0,99
	Afrontamiento	HSEREP2S002	-0,827	0,004	0,98	0,99
		HSEREP2S003	-1,174	0,004	0,97	0,95
		HSEREP2S004	-1,095	0,004	0,86	0,84
		HSEREP2S005	-0,835	0,004	0,85	0,84
		HSEREP2S006	-0,875	0,004	0,90	0,89
		HSEREP2S008	-0,741	0,004	1,09	1,11
	Control personal	HSEREP2S010	0,304	0,003	1,05	1,08
		HSEREP2S016	0,432	0,003	1,10	1,15
		HSEREP2S017	0,291	0,003	0,95	0,98
		HSEREP2S018	0,424	0,003	1,06	1,08
		HSEREP2S020	-0,037	0,003	1,29	1,31

4.8.2 Estructura de umbrales

Una característica adicional del modelo de escala de valoración de Rasch en relación con el modelo para datos dicotómicos es que, además de reportar las estimaciones de habilidad de las personas y de dificultad de cada ítem, provee una estructura de umbrales común para todos los ítems. Los umbrales de Andrich representan los puntos en que las categorías de respuesta adyacentes son igualmente probables. Así, cuando un estudiante tiene una medida promedio que excede el valor del umbral de dos categorías adyacentes, tiene una mayor probabilidad de ser observado en la categoría más alta (encima del umbral). Es importante que los umbrales estén ordenados de manera creciente, pues las observaciones en las categorías más altas deberían ser producidas por medidas más altas (Linacre, 1999). Esto quiere decir que las medidas promedio por categoría, para cada conjunto empírico de observaciones, deberían avanzar monótonicamente hacia arriba en la escala de valoración; de otro modo, su significado sería incierto y, consecuentemente, cualquier medida derivada sería de dudosa utilidad (Linacre, 1999). Asimismo, se recomienda que la distancia mínima entre los umbrales sea de por lo menos 1,4 logits, aunque esto no necesariamente es un requerimiento para construir medidas válidas e inferencialmente útiles a partir de las observaciones de una escala de valoración (Linacre, 1999).

Tabla 4.28 Estructura de umbrales de las escalas de HSE aplicadas en 2.º grado de secundaria

Escala	Primer umbral		Segundo umbral		Tercer umbral		Cuarto umbral	
	Medida	Error	Medida	Error	Medida	Error	Medida	Error
Toma de decisiones	-1,242	0,002	-1,039	0,002	0,957	0,002	1,324	-
Resolución de problemas	-1,758	0,002	-0,464	0,002	-0,053	0,002	2,275	-

4.8.3 Evidencias de validez vinculadas a la estructura interna

En la tabla 3, se presentan los resultados de este análisis. De estos, se concluye que todas las escalas y subescalas de HSE utilizadas en la EM 2022 cumplen con el requerimiento de unidimensionalidad, y que cada una de ellas mide fundamentalmente una sola variable latente.

Tabla 4.29 Análisis de unidimensionalidad de las medidas de HSE derivadas de las escalas aplicadas en 2.º grado de secundaria

Escala	Subescala	Medidas Rasch		ACPR	
		Autovalor	Varianza explicada (%)	Autovalor	Varianza explicada (%)
Toma de decisiones	Otros estilos	7,6	38,8	3,3	16,8
	Estilo vigilante	6,3	51,0	1,7	13,8
Resolución de problemas	Confianza	5,2	32,0	3,4	20,7
	Afrontamiento	8,1	40,2	2,2	10,7
	Control personal	5,5	47,7	1,4	12,5

4.8.4 Confiabilidad

La confiabilidad se calculó utilizando la estimación EAP (expected a posteriori) de las medidas de las personas. En relación con las escalas de HSE aplicadas en 2.º grado de secundaria, todas mostraron medidas con una confiabilidad por encima del punto de corte ($\leq 0,70$). Estos resultados aparecen en la tabla 4.30.

Tabla 4.30 Confiabilidad de las medidas de HSE derivadas de las escalas aplicadas en 2.º grado de secundaria

Escala	Subescala	Confiabilidad
Toma de decisiones	Otros estilos	0,85
	Estilo vigilante	0,71
Resolución de problemas	Confianza	0,82
	Afrontamiento	0,82
	Control personal	0,75

Procesamiento de datos de Factores Asociados

Capítulo 5

Los factores asociados son un conjunto de variables que inciden en los resultados de las evaluaciones de aprendizajes. Muchos de ellos son posibles objetos de intervención desde el aula, la escuela y el sistema educativo (Ravela, 2006). El término “asociado” sugiere la posibilidad de demostrar la existencia de asociaciones (si bien no necesariamente causales) entre estos factores y los resultados obtenidos por los estudiantes en evaluaciones estandarizadas.

En la ENLA 2023, la información de factores asociados se ha recogido a través de cuestionarios de autorreporte aplicados a los diferentes actores educativos (directores, docentes, estudiantes y familia) de cada uno de los grados evaluados (2.º y 4.º grados de primaria, 4.º grado de primaria EIB y 2.º grado de secundaria).

Cada uno de los cuestionarios aplicados está compuesto por un grupo de preguntas o ítems, los cuales son clasificados según el tipo de variable que busca recoger. Así, existen preguntas que buscan recoger información concreta de algún aspecto particular de la persona (por ejemplo: sexo, edad o años trabajando en la escuela). Por otro lado, también se consideran preguntas que están diseñadas para representar un constructo latente que no puede medirse directamente (por ejemplo: agotamiento emocional). Estas preguntas están conceptualmente relacionadas entre sí y se agrupan en una única medida compuesta que nos permita representar la variable latente a estudiar (Bandalos, 2018).

El siguiente capítulo busca brindar una descripción de las variables recogidas el presente año, así como los procedimientos realizados para su elaboración y todo lo vinculado con el análisis de dicha información.

5.1. Elaboración de las preguntas del cuestionario de factores asociados

El cuestionario es un instrumento que permite aproximarnos a una amplia gama de aspectos de interés que se encuentran alineados con el propósito de una investigación. Dentro de sus principales ventajas, se aprecia que es un instrumento que puede ser aplicado a una gran cantidad de personas de forma estandarizada, lo cual favorece la comparabilidad entre las respuestas de los participantes (Malhotra, 2016). Asimismo, permite recoger tanto información cuantitativa como cualitativa con relativa facilidad (Malhotra, 2016; Patel & Joseph, 2016).

Para el caso de factores asociados, los cuestionarios tienen el objetivo de contextualizar los resultados de aprendizaje en todos los grados y áreas evaluadas.

Con el objetivo de asegurar la calidad del recojo de la información, de tal manera que sea sensible a las características de los participantes y pueda responder correctamente a los objetivos y preguntas de investigación, es importante considerar su correcta elaboración y validación. En ese sentido, la presente sección detalla el proceso realizado para la preparación de los cuestionarios.

5.1.1. Algunos conceptos claves

Antes de comenzar con la descripción del procedimiento en sí mismo, es importante mencionar algunos términos que serán usados en esta sección: indicadores, índices, escalas y constructos. Al hablar de indicador, se hace referencia a una característica específica que se puede medir a través de una variable o ítem utilizado. Los indicadores son elementos básicos, usualmente observables o medibles que, en suma, pueden componer una escala o un índice (Bandalos, 2018). Algunos ejemplos de indicadores son los siguientes: número de estudiantes por aula, principal material con el que están construidas las paredes de una vivienda, nivel educativo más alto de un docente, frecuencia con la que un estudiante piensa en realizar sus tareas de Ciencia y Tecnología, etc.

Una escala se refiere a un conjunto de preguntas, ítems o indicadores diseñados para medir una variable específica. Los ítems en una escala suelen estar relacionados conceptualmente entre sí y se agrupan en una única medida compuesta (Bandalos, 2018). Usualmente, las escalas sirven para la medición de constructos, los cuales hacen referencia a un conjunto de construcciones teóricas o procesos mentales que se comprenden como propiedades subyacentes, que no pueden observarse de manera directa, y que permiten explicar ciertos comportamientos de una persona (DeVellis & Thorpe, 2021; Leppink, 2019). Algunos constructos son relativamente sencillos de medir, como la simpatía por un partido político, mientras que otros son más complejos, como el aprendizaje autorregulado. Para ello, en primer lugar, es importante contar con una clara conceptualización o definición teórica del constructo, la cual derivará en una definición operacional que, a su vez, se traducirá en indicadores. Cabe señalar que el nivel de complejidad de un constructo se traducirá en una mayor complejidad en su operacionalización; es decir, en el número o extensión de las preguntas, ítems o estímulos.

Por último, un índice es una construcción más amplia, que abarca múltiples variables, indicadores o aspectos que se relacionan entre sí. Estos se crean combinando varios indicadores o escalas para obtener una medida compuesta que representa una característica más amplia. Los índices pueden ser útiles cuando se desea capturar una idea muy compleja o multidimensional y que no puede abordarse con indicadores aislados o con una sola escala (Babbie, 2010). Un ejemplo de índice construido por los especialistas del equipo de Factores Asociados en la UMC es el índice socioeconómico.

5.1.2. Procedimiento para la elaboración de cuestionarios

Sobre el proceso de construcción de escalas, Furr (2013) enumera cuatro fases: 1) articulación del constructo y el contexto, 2) elaboración de ítems y escalas de respuesta, 3) obtención de los datos y 4) análisis psicométrico. En la presente sección se desarrollarán los puntos principales de las dos primeras fases.

A. Articulación del constructo y el contexto

El proceso para seleccionar los factores a priorizar comienza con la identificación de los grados y áreas evaluadas en cada año, así como los énfasis que se realizarán en cada uno de estos. Por ejemplo, para la ENLA 2023 se aplicó en 2.º grado de secundaria la prueba de Ciencias Sociales, lo que implica la elaboración de instrumentos vinculados con dicha área, así como la identificación de variables de factores asociados a las que se buscaría realizar seguimiento. Además de ello, se debe considerar que cada año existen diversas demandas de información debido a la coyuntura específica del país (por ejemplo: indagar en los cuestionarios sobre la educación remota debido al Covid-19), por lo que los cuestionarios de factores asociados también deben responder a estas necesidades de información.

Una vez definidas las áreas a evaluar en la ENLA 2023, se procedió a realizar una primera propuesta a partir de la revisión teórica y de los estudios empíricos sobre factores asociados a los logros de aprendizaje de las competencias a evaluar. La propuesta elaborada consideró lo siguiente:

- a. Enfatizar aquellos aspectos que requieren la atención e intervención de las distintas instancias del sistema educativo; por ejemplo: el trabajo pedagógico colegiado, las características del liderazgo directivo, el trabajo intra institucional e interinstitucional de las IE, aspectos referidos a la implementación del currículo traducidos en prácticas pedagógicas, creencias docentes, oportunidades de aprendizaje, etc.
- b. Elaborar una propuesta a partir de la información de factores asociados, recogida previamente para las cohortes evaluadas, e identificar la relevancia de hacer seguimiento a alguna de ellas.
- c. Dialogar preliminarmente con los subcoordinadores del equipo de Evaluación de la UMC con el fin de identificar temas de relevancia para comprender los logros de aprendizaje.
- d. Mapear los cambios más recientes en el sistema educativo, especialmente aquellos que pudieran haber afectado las prácticas escolares a nivel nacional (por ejemplo: compra de tecnologías para IE rurales).
- e. Plantear preguntas de investigación que puedan ser respondidas a través de los diferentes cuestionarios aplicados.

- f. Organizar los temas o ejes temáticos en categorías de análisis. Así, en la ENLA 2023 se emplearon las siguientes categorías: a) antecedentes del estudiante y su hogar; b) motivación, creencias, actitudes, sentimientos y comportamientos de los estudiantes; c) creencias, prácticas pedagógicas y oportunidades de aprendizaje; d) prácticas y políticas escolares, y antecedentes del docente y del director; y, e) políticas y prácticas a nivel de sistema educativo.

Tanto la revisión bibliográfica como la organización de la propuesta fueron los primeros pasos para tener mayor claridad conceptual con respecto a las diferentes variables que se evaluaron en los cuestionarios.

Asimismo, Furr (2013) señala la importancia de definir algunos aspectos de la naturaleza del constructo. Por ejemplo: la importancia de identificar si el constructo es más cercano a una actitud, una percepción, una atribución, una disposición, una respuesta emocional o una respuesta conductual. Por otro lado, es importante identificar las diferencias y similitudes con otros constructos relevantes, y si el constructo tiene subdimensiones que tendrían que ser diferenciadas.

Finalmente, la información recogida fue sistematizada en una propuesta preliminar, empleada luego como punto de partida para la discusión y elaboración de una propuesta consensuada por los equipos de la UMC y, posteriormente, aprobada por la jefatura.

B. Elaboración de ítems y escalas de respuesta

Para la redacción de los ítems, se tomaron en cuenta los criterios de Stehr et al. (2003), así como de Malhotra (2016), los cuales señalan que estos deben incluir una idea clara y sin ambigüedad. Asimismo, se usó un vocabulario simple, evitando el lenguaje técnico, las abreviaturas, las negaciones y las dobles negaciones, por lo que los ítems debían ser concretos respecto al tiempo y a los eventos de referencia. Asimismo, para evitar el sesgo en la formulación de preguntas o ítems se prescindió de colocar palabras o textos que direccionen la respuesta de los participantes (Malhotra, 2016).

Con respecto a la longitud de las escalas y la cantidad de ítems por cada una de estas, se esperaba contar con un mínimo de cuatro ítems para el caso de constructos unidimensionales, ya que esta cantidad produce soluciones convergentes en el análisis factorial confirmatorio. Sin embargo, dado que se suelen eliminar ítems problemáticos, es recomendable incluir un ítem adicional, por lo que un mínimo de cinco ítems en una escala fue considerado lo ideal¹. Más allá de esto, el número máximo de ítems por escala dependerá de la complejidad de la variable que se

¹ Para constructos con varias dimensiones el mínimo número que se acepta para formar una dimensión es tres debido a que en este tipo de modelo se pueden obtener soluciones que converjan.

mida, ya que se necesitará un mayor número de ítems para captar la riqueza de una variable multidimensional; no obstante, ello debe equilibrarse con la brevedad, para maximizar las tasas de respuesta (Furr, 2013; Robinson, 2018).

C. Validación de la selección de instrumentos, temas y contenidos con stakeholders Minedu.

Una vez elaborada la propuesta final de escalas e ítems, estos pasaron a través de un proceso de validación por parte de otros especialistas del Minedu para comprobar su pertinencia. Los especialistas han sido seleccionados según el ámbito educativo de acción; por ejemplo: si se recoge información pedagógica, se busca dialogar la propuesta con la Dirección de Formación Docente en Servicio (DIFODS).

Asimismo, para el caso de nuevas escalas, se suele usar el método de la **entrevista cognitiva** como una forma alternativa y funcional para validar y mejorar los contenidos de los cuestionarios. Este método se emplea a menudo para probar las preguntas y determinar cómo deben modificarse antes de aplicar un instrumento, haciéndolo más comprensible y fácil de responder. La entrevista cognitiva aborda el proceso de responder a un instrumento desde el *cognition and survey methodology* (CASM, por sus siglas en inglés), por lo que puntualiza cuatro procesos o etapas cognitivas: 1) la comprensión de la pregunta; 2) la recuperación de la información necesaria para responder a la pregunta (memoria); 3) los procesos de decisión o estimación, especialmente en relación con la idoneidad de la respuesta o la amenaza potencial que puede plantear el contenido sensible o la deseabilidad social (juicio); y 4) el proceso de respuesta en que el encuestado produce una respuesta que satisface los requisitos de la tarea (p. ej.: hacer coincidir una respuesta generada internamente con una respuesta de la encuesta [Willis, 2008]).

Una vez realizados estos procesos, las escalas se encontraron listas para ser diagramadas y aplicadas.

D. Análisis de las escalas de factores asociados

Después del proceso de aplicación de los cuestionarios, se inició el análisis de las bases de datos. Como primer paso, las bases de datos de factores asociados pasaron por un proceso de revisión, en el cual se recodificaron todas las opciones de respuesta en formato numérico, se asignó las etiquetas a cada valor y se añadió información adicional sobre la pregunta (lo que comúnmente se denomina metadata). Una vez que se contó con las bases completamente revisadas, se procedió con el análisis, comenzando desde los aspectos más sencillos (revisión de casos perdidos, descriptivos) hasta los más complejos (análisis factorial confirmatorio, análisis multinivel, entre otros).

5.2. Elaboración de análisis descriptivo e imputación de datos

El análisis de descriptivo permitió la representación gráfica de las distribuciones de respuesta de cada ítem que compone los cuestionarios. Al haber sido la mayoría ítems categóricos, lo que se presentó fueron los porcentajes de respuesta en cada una de las opciones de marcado. Estos reportes son de utilidad para identificar patrones de respuesta en las diferentes preguntas.

El análisis descriptivo también ayuda a identificar el porcentaje de valores perdidos en cada pregunta, con el fin de detectar posibles problemas para su posterior uso. Cuando los valores perdidos superan un porcentaje mayor al 10 %, se considera revisar dicho ítem e identificar si existen patrones en el marcado que podrían explicar dicha distribución (por ejemplo: si es que hay un mayor porcentaje de valores perdidos en estudiantes mujeres que en estudiantes hombres). De encontrar algún patrón irregular, se establece una alerta para el uso de dichas preguntas. Los ítems que componen preguntas simples no pasan por ningún proceso adicional; sin embargo, las preguntas que componen las escalas pasan por un proceso de imputación que es señalado a continuación

La imputación de datos se usa únicamente para la estimación de los parámetros de las escalas e índices². Esto es debido a que el método de tratamiento de casos con valores perdidos que usa R en varios de sus paquetes (por ejemplo: *lavaan* o *psych*) es el de eliminación de la lista (*listwise deletion*, por sus siglas en inglés). Este método solo considera, para la estimación, a los participantes que tengan respuestas en todos los ítems que componen la escala. Esto implica que, si hay una escala conformada por 10 ítems y el participante presenta un valor perdido en al menos uno de estos, queda eliminado para dicha estimación. La imputación es de tipo simple y se usa el paquete *mice* de R con la semilla “2022” y con el método *predictive mean matching*³. Una vez que se cuenta con toda la información señalada, se procede con los posteriores análisis.

5.3. Validez y confiabilidad de los índices y escalas de factores asociados

Luego de la realización del reporte de análisis descriptivo, se procedió con la obtención de evidencias de validez y confiabilidad, así como de los puntajes de las escalas que se elaboraron.

Los análisis estadísticos que realiza la UMC a sus escalas dependen del enfoque que se ha utilizado para conceptualizar cada una de estas, además de la dirección

¹ Las respuestas imputadas son usadas únicamente para la estimación de puntajes de las escalas e índices. De esta manera, las bases finales cuentan con las respuestas tal como fueron marcadas por los participantes.

² Pese a que el método *predictive mean matching* se suele usar para variables continuas, pues arroja una “media” de los valores presentados, cuando todas las variables son discretas este método también arroja valores discretos.

entre los ítems que componen la escala (variables manifiestas) y la variable latente. En caso de que las respuestas a dichos ítems representen el efecto de la variable latente, se está frente a un modelo reflexivo (Bollen y Bauldry, 2011; Jarvis et al., 2003). Un ejemplo de este tipo de variables es la escala de ansiedad a la matemática, en la cual los ítems representan acciones que se dan como consecuencia de tener un alto o bajo nivel en dicha variable latente. Por el contrario, si las variables manifiestas son la causa de la variable latente, estaríamos frente a un modelo formativo, como es el caso del índice socioeconómico (ISE), en el cual los ítems son un componente que conforma la variable latente del ISE.

Dadas las diferentes formas en que se podría conceptualizar una escala, estas son analizadas mediante diferentes métodos. Para el caso de los constructos reflexivos, se realiza un análisis factorial confirmatorio (AFC), con el fin de confirmar la estructura y dimensionalidad de las escalas. En caso de que el constructo sea de tipo formativo, las estimaciones se realizan mediante un análisis de componentes principales (ACP). Asimismo, debido a que todos los ítems que se utilizan para las escalas (independientemente de si son constructos formativos o reflexivos) son categóricas, se realizan los ajustes necesarios en las estimaciones para trabajar con ellos. Los análisis son realizados a través de diversos paquetes del lenguaje de programación R.

5.3.1. Evidencias de validez basadas en la estructura interna

a. Para constructos reflexivos

Tal como se señaló anteriormente, las evidencias de validez para constructos reflexivos son recolectadas a través del AFC. Al utilizarlo, se necesita testear el modelo teórico de la dimensionalidad de los ítems por medio de datos empíricos⁴. Este modelo vincula las variables observables (ítems) con la variable latente a través de la siguiente fórmula:

$$\gamma = \Lambda_y \eta + \varepsilon$$

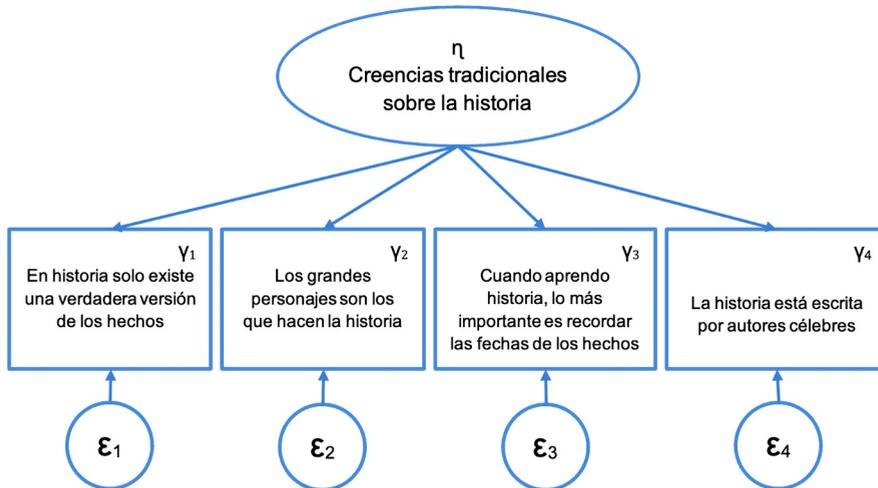
Dicha fórmula establece la relación entre las respuestas de los ítems (γ) a través de la variable latente (η) con una ecuación muy similar a la de una regresión. Este modelo considera la matriz de cargas factoriales (Λ), que vendrían a ser los coeficientes de la regresión, y un vector de residuos ε , que vendrían a ser las desviaciones de cada observación desde la línea de regresión proyectada.

A continuación, se muestra una representación gráfica de la ecuación descrita anteriormente mediante un ejemplo con una de las escalas de la ENLA 2023 aplicada a estudiantes (figura 5.1).

⁴ Como se vio en la sección anterior, todas las escalas elaboradas por el equipo de factores asociados tienen un objetivo específico para la medición. Además, se elaboran a partir de un proceso de revisión bibliográfica y de validación por otros expertos, por lo que se decide hacer dicho CFA. En algunos casos, también se realizan análisis factoriales exploratorios (EFA, por sus siglas en inglés), pero para escalas nuevas en que no se tenga mucha claridad conceptual.

La variable latente en el ejemplo (creencias tradicionales sobre la historia) es comúnmente representada con una elipse. Los ítems o variables observables son representadas con cuadrados y los residuos, con círculos. Las flechas parten desde

Figura 5.1. Representación gráfica de un constructo reflexivo



la variable latente hacia las variables observables, debido a que es un constructo reflexivo.

A partir de una estructura teórica, se estiman parámetros que generan una matriz reproducida, la cual se compara con la empírica para evaluar su ajuste. Uno de los primeros métodos de estimación que se utilizaron para minimizar la discrepancia entre la matriz predicha y la observada suele ser el de máxima verosimilitud (ML por sus siglas en inglés). Sin embargo, el uso de este estimador trata a las variables observables como si fueran medidas continuas; por lo tanto, no resultaría apropiado, pues todos los ítems utilizados son de tipo categórico. Además, este método de estimación precisa de normalidad multivariada, implicando que la distribución en conjunto de ítems que forman el constructo latente sea normal, lo que se puede incumplir en las escalas de tipo Likert (Flora y Curran, 2004). Por ello, el estimador utilizado y que se adecúa más a este tipo de variables es el de Mínimos Cuadrados Ponderados por Media y Varianza (WLSMV, por sus siglas en inglés; Muthén et al., 1997; Flora y Curran, 2004; DiStefano y Morgan, 2014), el cual usa matrices de correlaciones policóricas con los ítems para cada indicador.

La bondad de ajuste de los modelos es examinada a través de los siguientes indicadores: Índice de Ajuste Comparativo (CFI), Índice de Tucker-Lewis (TLI), Error

Cuadrático Medio de Aproximación (RMSEA) y el Residual Estandarizado de la Raíz Cuadrada Media (SRMR) (Marsh et al., 2004; Chen et al., 2009; Lai y Green, 2016). A pesar de que no existe un consenso respecto de qué es un buen ajuste, hay varios autores que coinciden en algunos puntos de corte. Así, a partir de la lectura de estos textos, el equipo de Factores Asociados considera que valores por encima de ,95 de CFI y TLI nos hablarían de un buen ajuste en el modelo (Hu y Bentler, 1999; Marsh et al., 2004; Lai y Green, 2016). De otro lado, valores inferiores a ,10 de RMSEA y SRMR indicarían valores marginalmente satisfactorios (MacCallum et al., 1996; Rutkowski y Svetina, 2013). Además de estos índices de ajuste, se tomaron en cuenta las cargas factoriales estandarizadas de los ítems, pues estas brindan información adicional de la educación del modelo a los datos del cuestionario. En este sentido, se consideran como cargas “adecuadas” aquellas que superen 0,4 como valor (Ford et al., 1986).

Todas estas estimaciones se realizaron utilizando el paquete lavaan (Rosseel, 2012) del lenguaje R (<https://cran.r-project.org/>). Los resultados de las escalas se muestran en la tabla 5.1.

Tabla 5.1. Indicadores de ajuste para constructos reflexivos según grado y actor

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR	RMSEA
2.º grado de primaria	Director	DIRPGEN_SITPROB	Situaciones problemáticas en la escuela	0,99	0,98	0,059	0,061
		DIRPGEN_EDUIN	Actitudes hacia la educación inclusiva	1	1	0,03	0,049
	Docente	DOC2PLEC_ESTLEC	Estrategias para la enseñanza de la comprensión lectora	0,99	0,98	0,071	0,08
		DOC2PMAT_ESTMAT	Estrategias para la enseñanza de la matemática	0,99	0,98	0,069	0,078
		DOC2PMAT_CONF	Confianza de los docentes para enseñar matemática	0,99	0,99	0,059	0,064
		DOC2PMAT_DIFMAT	Dificultades para el desarrollo de la capacidad de Resuelve problemas de cantidad	0,99	0,99	0,045	0,062
		DOC2PGEN_PROBIE	Percepción de situaciones problemáticas en la escuela	0,99	0,99	0,062	0,065
		DOC2PMAT_AUTMAT	Autoeficacia hacia la matemática	1	1	0,022	0,056

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR	RMSEA
2.º grado de primaria	Familia	FAM2PGEN_BIENEST	Malestar socioafectivo del estudiante	1	1	0,013	0,021
		FAM2PLEC_LACTTEMP	Actividades tempranas de literacidad	0,98	0,97	0,06	0,099
		FAM2PMAT_MACTTEMP	Actividades tempranas de aritmética	0,99	0,98	0,043	0,073
		FAM2PGEN_INVHOG	Involucramiento parental en el hogar	1	1	0,029	0,057
		FAM2PGEN_INVESC	Involucramiento parental: interacción con los docentes	0,99	0,99	0,042	0,063
		FAM2PLEC_INTLEC	Interés por la lectura	0,99	0,97	0,047	0,134
		FAM2PGEN_EDUIN	Actitudes hacia la educación inclusiva	1	1	0,035	0,053
4.º grado de primaria	Docente de Comunicación	DOC4PGEN_PROBIE	Percepción de situaciones problemáticas en la escuela	0,99	0,99	0,047	0,054
		DOC4PLEC_ESTLEC	Estrategias para la enseñanza de la comprensión lectora	0,99	0,99	0,054	0,068
		DOC4PLEC_MOTLEC	Prácticas para la motivación de la lectura	0,97	0,96	0,07	0,082
		DOC4PGEN_EDUIN	Actitudes hacia la educación inclusiva	1	1	0,024	0,039
		DOC4PGEN_EVFORM	Creencias sobre la evaluación formativa	0,96	0,95	0,064	0,094
	Docente de Matemática	DOC4PGEN_PROBIE	Percepción de situaciones problemáticas en la escuela	0,99	0,99	0,049	0,056
		DOC2PMAT_CONF	Confianza de los docentes para enseñar matemática	0,99	0,99	0,055	0,077
		DOC4PMAT_NAT	Creencias sobre la naturaleza de las matemáticas	0,95	0,93	0,085	0,119
		DOC4PMAT_APR	Creencias sobre la enseñanza de la matemática	0,95	0,94	0,066	0,088
		DOC4PGEN_EDUIN	Actitudes hacia la educación inclusiva	1	1	0,024	0,041

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR	RMSEA
4.º grado de primaria	Familia	FAM4PGEN_BIENEST	Malestar socioafectivo del estudiante	1	0,99	0,017	0,028
		FAM4PGEN_INVHOG	Involucramiento parental en el hogar	1	1	0,024	0,043
		FAM4PGEN_INVESC	Involucramiento parental: interacción con los docentes	0,99	0,99	0,038	0,055
		FAM4PLEC_INTLEC	Interés por la lectura	0,99	0,99	0,035	0,077
		FAM4PGEN_ESTDES	Estereotipos de género en el desempeño	1	1	0,027	0,101
		FAM4PGEN_ESTEMO	Estereotipos de género en las emociones	1	1	0,023	0,079
		FAM2PGEN_EDUIN	Actitudes hacia la educación inclusiva	1	0,99	0,036	0,055
4.º grado de primaria EIB	Director	DIRPEIB_USOTIEM	Uso de tiempo del director	0,99	0,99	0,049	0,074
		DIRPEIB_ROLIDER	Rol del líder de la comunidad en la IE	0,99	0,99	0,054	0,097
		DIRPEIB_INVFAM	Involucramiento parental en la escuela	1	1	0,019	0,05
		DIRPEIB_VALSOC	Valoración social de la LO	0,99	0,99	0,04	0,056
		DIRPEIB_PERCRER	Percepción de las redes educativas	1	1	0,018	0,111
4.º grado de primaria EIB	Docente	DOC4PEIB_BIENES	Bienestar subjetivo	1	0,99	0,029	0,065
		DOC4PEIB_RECCOMUN	Percepción sobre la recepción de la comunidad	0,99	0,98	0,05	0,081
		DOC4PEIB_PERGESDIR	Percepción de la gestión del director	1	1	0,043	0,093
		DOC4PEIB_PARGIA	Participación en los GIA	1	1	0,021	0,068
		DOC4PEIB_ELECTXTL2	Uso de estrategias para la elección de textos en Castellano	0,99	0,99	0,041	0,056
		DOC4PEIB_ELECTXTL0	Uso de estrategias para la elección de textos en LO	0,99	0,99	0,041	0,072

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR	RMSEA	
4.º grado de primaria EIB	Docente	DOC4PEIB_LECCOMAP	Creencias sobre la lectura	1	0,99	0,02	0,035	
		DOC4PEIB_MOTDOC	Motivación para la docencia	0,99	0,99	0,055	0,063	
	Familia	FAM4P_ACOMPFAM	Acompañamiento familiar en los aprendizajes	0,99	0,99	0,037	0,027	
		FAM4P_CONESFAM	Convocatoria de la escuela a la familia	0,99	0,98	0,031	0,029	
		FAM4P_INVFAM	Involucramiento parental en la escuela	0,99	0,97	0,053	0,064	
		FAM4P_VALSOC	Valoración social de la LO	0,97	0,96	0,063	0,039	
2.º grado de secundaria	Docente tutor	DOC2SHSE_CRAS	Creencias sobre el aprendizaje socioemocional	0,99	0,98	0,08	0,098	
		DOC2SHSE_INT	Interés por el bienestar de los estudiantes	0,98	0,97	0,054	0,063	
		DOC2SHSE_EMP	Empatía cognitiva docente	0,99	0,99	0,039	0,06	
		DOC2SHSE_AUT	Autoeficacia docente	0,99	0,99	0,053	0,086	
		DOC2SHSE_RDE	Relación docente-estudiante	1	0,99	0,035	0,065	
			DOC2SHSE_RE	Regulación emocional	0,99	0,99	0,05	0,068
	Docente de Ciencias Sociales		DOC2SCSS_CIUAC	Prácticas que promueven la Ciudadanía activa	0,97	0,96	0,071	0,078
			DOC2SCSS_ACTEVA	Actividades de evaluación	0,95	0,94	0,097	0,082
			DOC2SCSS_PROBIE	Percepción de situaciones problemáticas en la escuela	0,99	0,99	0,053	0,076
			DOC2SCSS_EDUIN	Actitudes hacia la educación inclusiva	1	1	0,023	0,036
		DOC2SCSS_EVFORM	Creencias sobre la evaluación formativa	0,97	0,96	0,056	0,083	

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR	RMSEA
2.º grado de secundaria	Docente de Comunicación	DOC2SLEC_ESTRLEC	Enseñanza sobre estrategias de lectura	0,99	0,98	0,05	0,054
		DOC2SLEC_ESTRCOMPLEC	Prácticas pedagógicas en Comunicación con énfasis en la repetición	0,99	0,99	0,053	0,065
		DOC2SLEC_AUTLEC	Autoeficacia de docente para la enseñanza de la lectura	1	0,99	0,033	0,055
		DOC2SLEC_MOTLEC	Motivación a la enseñanza de lectura	0,98	0,97	0,058	0,07
		DOC2SLEC_PROBIE	Percepción de situaciones problemáticas en la escuela	0,99	0,99	0,051	0,071
		DOC2SLEC_EDUIN	Actitudes hacia la educación inclusiva	1	1	0,028	0,043
		DOC2SLEC_EVFORM	Creencias sobre la evaluación formativa	0,96	0,95	0,066	0,098
	Docente de Matemática	DOC2SMAT_PROBIE	Percepción de situaciones problemáticas en la escuela	0,98	0,98	0,062	0,074
		DOC2SMAT_APR	Creencias sobre la enseñanza de la matemática	0,95	0,93	0,062	0,07
		DOC2SMAT_NAT	Creencias sobre la naturaleza de las matemáticas	0,95	0,93	0,088	0,106
		DOC2SMAT_HABIN	Creencias sobre la matemática como habilidad innata	0,99	0,98	0,039	0,066
		DOC2SMAT_EDUIN	Actitudes hacia la educación inclusiva	1	1	0,028	0,045
		DOC2SMAT_EVFORM	Creencias sobre la evaluación formativa	0,96	0,95	0,062	0,094
	Estudiante	EST2SGEN_INVHOG	Involucramiento parental en el hogar	1	1	0,03	0,057
		EST2SGEN_SENPER	Sentido de pertenencia	1	1	0,021	0,035
EST2SGEN_PERNORM		Percepción de las normas y disciplina	1	1	0,015	0,037	

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR	RMSEA
2.º grado de secundaria	Estudiante	EST2SGEN_INTEXT	Interés por los textos en la escuela	1	1	0,023	0,056
		EST2SGEN_MOTL	Motivación hacia la lectura: tiempo libre	0,99	0,99	0,07	0,096
		EST2SMAT_CLIMER	Clima del error en Matemática	0,95	0,93	0,066	0,089
		EST2SMAT_EVFORM	Prácticas de evaluación formativa en Matemática	0,99	0,98	0,051	0,092
		EST2SMAT_ANSIE	Ansiedad ante la matemática	1	1	0,029	0,081
		EST2SCSS_DISCU	Discusión en clases de CCSS	1	1	0,032	0,075
		EST2SCSS_CREHIST	Pensamiento histórico del estudiante	0,97	0,96	0,048	0,065
		EST2SCSS_CIUAC	Ciudadanía activa	0,97	0,96	0,072	0,089
		EST2SCSS_ACTEVA	Actividades de evaluación en CCSS	0,99	0,99	0,032	0,055
		EST2SLEC_ACTCOG	Activación cognitiva en clases de Comunicación	0,99	0,99	0,04	0,074
EST2SLEC_MOTL	Motivación hacia la lectura: en la escuela	0,99	0,99	0,057	0,086		
EST2SHSE_CSE	Clima social escolar	0,99	0,99	0,045	0,065		
EST2SHSE_EAA	Estrés académico adolescente	0,98	0,97	0,063	0,09		

b. Para constructos formativos

El análisis de constructos formativos fue realizado por medio del ACP. El objetivo de este análisis es formar nuevas variables, llamadas *componentes principales*, a partir de un conjunto de variables de entrada posiblemente correlacionadas (Mair, 2018). Sea X una matriz conteniendo m indicadores, ACP extrae nuevas variables independientes entre sí, c_1, c_2, \dots, c_m . Para el primer componente principal, la ecuación que relaciona c con los indicadores es:

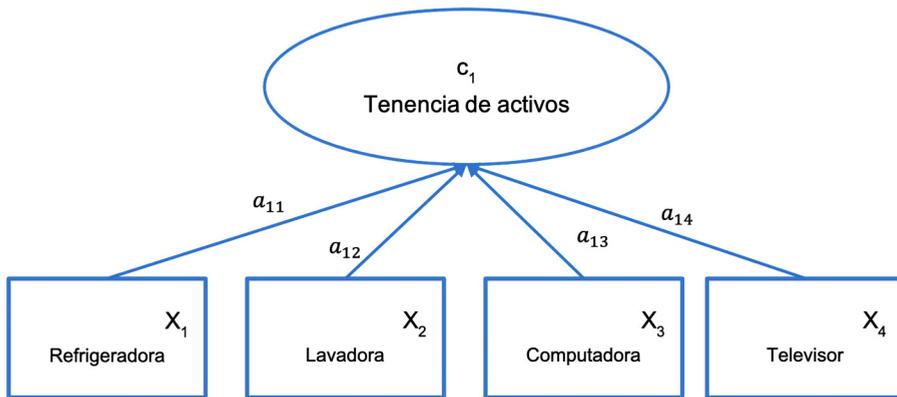
$$c_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m$$

Donde a representa las cargas. El primer componente principal c_1 capta la mayor variabilidad de los datos. El segundo componente principal c_2 , debe ser no correlacionado con el primero y debe captar la mayor variabilidad que no ha sido expresada por el primero. c_3 debe ser no correlacionado con los dos primeros y debe captar la mayor variabilidad que no ha sido expresada por los primeros. Y así sucesivamente.

Para que c_m capte la mayor variabilidad de los datos, los coeficientes (cargas) de la combinación lineal que la forman deben ser las coordenadas del autovector a_m de la matriz de covarianzas de los indicadores originales, lo que corresponde al autovalor λ_m . Como propiedad, el autovalor λ_m corresponde a la varianza de cada componente principal. Si los k primeros autovalores representan una gran parte de la suma de todas las varianzas, será suficiente usar tan solo los k primeros componentes principales en lugar de las m variables originales para tener una gran parte de la información. Para el caso de la ENLA 2023, se esperó que el porcentaje de varianza explicada del primer componente no sea menor a 50 % e, idealmente mayor a 70 % (Jolliffe y Cadima, 2017).

Una representación gráfica del modelo formativo puede verse en la siguiente figura:

Figura 5.2. Representación gráfica de un constructo formativo



Estos indicadores son calculados a través de matrices de correlaciones policóricas con los ítems para cada indicador. Las correlaciones al interior de cada dimensión en la ENLA 2023 fueron positivas. Todas estas estimaciones se realizaron utilizando el paquete psych (Revelle, 2024) del lenguaje R (<https://cran.r-project.org/>) y se observan en la tabla 5.2.

Tabla 5.2. Indicadores de ajuste para constructos formativos según grado y actor

Grado	Actor	Código de constructo	Nombre	Varianza explicada
2.º grado de primaria	Director	DIRPGEN_SITDISR	Situaciones disruptivas en la escuela	52,32
	Docente tutor	DOC2SHSE_INV	Involucramiento institucional para el desarrollo de las HSE	64,46
		DOC2SHSE_DIS	Disrupción en el aula	62,53
2.º grado de secundaria	Docente de Comunicación	DOC2SLEC_PREPCOM	Preparación del docente para la enseñanza de la Lectura	73,3
		DOC2SLEC_TIPFORTE	Uso de tipos y formatos textuales	53,47
	Docente de Matemática	DOC2SMAT_ANSIEST	Ansiedad de los estudiantes hacia la matemática	67,7
4.º grado de primaria EIB	Docente de Comunicación	DOC4PLEC_TIPFORM	Uso de tipos y formatos textuales	49,43
		DIRPEIB_NEDCAP	Necesidades de capacitación	51,88
		DIRPEIB_INFRAIE	Infraestructura escolar	53,16
		DIRPEIB_DISMATLO	Disponibilidad de materiales: Material de lectura en LO	61,49
		DIRPEIB_DISMATL2	Disponibilidad de materiales: Material de lectura en L2	64,45
		DIRPEIB_ACCESTIC	Acceso a TIC y cobertura de señal en la IE	56,16
		DIRPEIB_MOTINASIST	Motivos de inasistencia a clases	58,41
	DIRPEIB_APOYIGED	Apoyo de las IGED a la IE	61,86	
	Docente	DOC4PEIB_NEDCAP	Necesidades de capacitación	53,61
		DOC4PEIB_VIVCARACT	Condiciones de vivienda del docente	82,98

5.3.2. Evidencias de Confiabilidad

Para el caso de las escalas de factores asociados, se usa el índice de confiabilidad omega (ω) de McDonald. Este coeficiente atiende los supuestos de unidimensionalidad de la escala y supone algunas ventajas sobre el tradicional alfa de Cronbach. En primer lugar, no supone la tau-equivalencia, la cual rara vez se da en la práctica ya que los ítems suelen tener diferentes cargas factoriales; en segundo lugar, no asume que las variables son continuas cuando en realidad cada ítem es una variable categórica, como fue mencionado en el apartado acerca de los AFC. La fórmula general de ω es la siguiente:

$$p = \frac{\left(\sum_i \lambda_i \right)^2}{\left(\sum_i \lambda_i \right)^2 + \sum_i \theta_i}$$

Donde λ son las cargas factoriales de cada indicador y θ la varianza residual de cada indicador. Para el cálculo de ω , se utilizó la función *compReISEM* del paquete *semTools* del software estadístico R (Jorgensen et al., 2022). Los resultados de confiabilidad aparecen en la tabla 5.3.

Tabla 5.3. Confiabilidad de las escalas según grado y actor

Grado	Actor	Código de constructo	Nombre general	Sub escala	omega	
2.º grado de primaria	Director	DIRPGEN_SITPROBD	Situaciones problemáticas en la escuela	Docentes	0,81	
		DIRPGEN_SITPROBE		Relacionadas al estudiante	0,903	
		DIRPGEN_SITPROBR		Falta de recursos	0,821	
		DIRPGEN_DEDUINAF		Afectiva	0,83	
		DIRPGEN_DEDUINCD	Actitudes hacia la educación inclusiva	Conductual	0,894	
		DIRPGEN_DEDUINCG		Cognitiva	0,82	
	Docente	DOC2PLEC_ESTLECD			Decodificación	0,623
		DOC2PLEC_ESTLECRE	Estrategias para la enseñanza de la comprensión lectora		Repetición	0,891
		DOC2PLEC_ESTLECSI			Significado	0,759
		DOC2PMAT_ESTMATCO	Estrategias para la enseñanza de la matemática		Constructivista	0,81
		DOC2PMAT_ESTMATTR			Tradicional	0,805
		DOC2PMAT_CONF	Confianza de los docentes para enseñar matemática			0,899
		DOC2PMAT_DIFMAT	Dificultades para el desarrollo de la capacidad de Resuelve problemas de cantidad			0,859
		DOC2PGEN_PROBIEE	Percepción de situaciones problemáticas en la escuela		Relacionadas al estudiante	0,899
DOC2PGEN_PROBIER		Falta de recursos		0,797		
DOC2PMAT_AUTMAT	Autoeficacia hacia la matemática			0,79		

Grado	Actor	Código de constructo	Nombre general	Sub escala	omega
2.º grado de primaria	Familia	FAM2PGEN_BIENEST	Malestar socioafectivo del estudiante		0,682
		FAM2PLEC_LACTTEMP	Actividades tempranas de literacidad		0,829
		FAM2PMAT_MACTTEMP	Actividades tempranas de aritmética		0,818
		FAM2PGEN_INVHOG	Involucramiento parental en el hogar		0,884
		FAM2PGEN_INVESC_ACT	Involucramiento parental	Participación en las actividades de la escuela	0,863
		FAM2PGEN_INVESC_DOC		Interacción con los docentes	0,785
		FAM2PLEC_INTLEC	Interés por la lectura		0,789
		FAM2PGEN_EDUINAF		Afectiva	0,796
		FAM2PGEN_EDUINCD	Actitudes hacia la educación inclusiva	Conductual	0,843
		FAM2PGEN_EDUINCG		Cognitiva	0,861
4.º grado de primaria	Docente de Comunicación	DOC4PGEN_PROBIEE	Percepción de situaciones problemáticas en la escuela	Relacionadas al estudiante	0,887
		DOC4PGEN_PROBIER		Falta de recursos	0,74
		DOC4PLEC_ESTLECDE		Decodificación	0,649
		DOC4PLEC_ESTLECRE	Estrategias para la enseñanza de la comprensión lectora	Repetición	0,876
		DOC4PLEC_ESTLECSI		Significado	0,776
		DOC4PLEC_MOTLEC	Prácticas para la motivación de la lectura		0,729
		DOC4PGEN_EDUINAF		Afectiva	0,854
		DOC4PGEN_EDUINCD	Actitudes hacia la educación inclusiva	Conductual	0,887
		DOC4PGEN_EDUINCG		Cognitiva	0,819

Grado	Actor	Código de constructo	Nombre general	Sub escala	omega
4.º grado de primaria	Docente de Comunicación	DOC4PGEN_EVFORMCR	Creencias sobre la evaluación formativa	Criterios erróneos de evaluación	0,797
		DOC4PGEN_EVFORMRE		Creencias erróneas sobre retroalimentación	0,66
	Docente de Matemática	DOC4PGEN_PROBIEE	Percepción de situaciones problemáticas en la escuela	Relacionadas al estudiante	0,889
		DOC4PGEN_PROBIER		Falta de recursos	0,744
		DOC2PMAT_CONF	Confianza de los docentes para enseñar matemática		0,915
		DOC4PMAT_NATPROC	Creencias sobre la naturaleza de las matemáticas	Proceso de búsqueda	0,752
		DOC4PMAT_NATREG		Reglas y procesos	0,763
		DOC4PMAT_APRACT	Creencias sobre la enseñanza de la matemática	Enfoque de aprendizaje activo	0,654
		DOC4PMAT_APRTRA		Enfoque tradicional	0,74
		DOC4PGEN_EDUINAF		Afectiva	0,861
		DOC4PGEN_EDUINCD	Actitudes hacia la educación inclusiva	Conductual	0,887
	DOC4PGEN_EDUINCG		Cognitiva	0,816	
	Familia	FAM4PGEN_BIENEST	Malestar socioafectivo del estudiante		0,632
		FAM4PGEN_INVHOG	Involucramiento parental en el hogar		0,866
		FAM4PGEN_INVESC_ACT	Involucramiento parental: interacción con los docentes		0,846
FAM4PGEN_INVESC_DOC				0,757	
FAM4PLEC_INTLEC		Interés por la lectura		0,778	
FAM4PGEN_ESTDES		Estereotipos de género en el desempeño		0,913	
FAM4PGEN_ESTEMO		Estereotipos de género en las emociones		0,839	

Grado	Actor	Código de constructo	Nombre general	Sub escala	omega
4.º grado de primaria	Familia	FAM2PGEN_EDUINAF		Afectiva	0,802
		FAM2PGEN_EDUINCD	Actitudes hacia la educación inclusiva	Conductual	0,826
		FAM2PGEN_EDUINCG		Cognitiva	0,845
4.º grado de primaria EIB	Director	DIRPEIB_TACTPED	Uso de tiempo del director	Actividades pedagógicas	0,824
		DIRPEIB_TADMIN		Actividades administrativas	0,74
		DIRPEIB_ROLIDER	Rol del líder de la comunidad en la IE		0,937
		DIRPEIB_INVFAM	Involucramiento parental en la escuela		0,82
		DIRPEIB_VALSOC	Valoración social de la LO		0,85
		DIRPEIB_PERCRER	Percepción de las redes educativas		0,916
	Docente	DOC4PEIB_BIENES	Bienestar subjetivo		0,807
		DOC4PEIB_RECCOMUN	Percepción sobre la recepción de la comunidad		0,81
		DOC4PEIB_PERGESDIR	Percepción de la gestión del director		0,926
		DOC4PEIB_PARGIA	Participación en los GIA		0,951
DOC4PEIB_ELECTXTL2		Uso de estrategias para la elección de textos en Castellano		0,767	
DOC4PEIB_ELECTXTLO		Uso de estrategias para la elección de textos en LO		0,841	
DOC4PEIB_LECTLN		Creencias sobre la lectura		0,742	
DOC4PEIB_LECCOMAP				0,589	
DOC4PEIB_UTPER	Motivación para la docencia	Utilidad personal	0,856		
DOC4PEIB_UTSOC		Utilidad social	0,735		

Grado	Actor	Código de constructo	Nombre general	Sub escala	omega	
4.º grado de primaria EIB	Familia	FAM4P_ACOMPFAM	Acompañamiento familiar en los aprendizajes		0,655	
		FAM4P_CONESFAM	Convocatoria de la escuela a la familia		0,497	
		FAM4P_INVFAM	Involucramiento parental en la escuela		0,711	
		FAM4P_VALSOC	Valoración social de la LO		0,653	
2.º grado de secundaria		DOC2SHSE_CRASCF		Confort	0,83	
		DOC2SHSE_CRASCL	Creencias sobre el aprendizaje socioemocional	Cultura	0,866	
		DOC2SHSE_CRASCM		Compromiso	0,865	
	Docente tutor	DOC2SHSE_INT	Interés por el bienestar de los estudiantes		0,698	
		DOC2SHSE_EMP	Empatía cognitiva docente		0,856	
		DOC2SHSE_AUT	Autoeficacia docente		0,917	
		DOC2SHSE_RDE	Relación docente-estudiante		0,793	
		DOC2SHSE_REEN		Expresión de emociones negativas	0,873	
		DOC2SHSE_REEP	Regulación emocional	Expresión de emociones positivas	0,788	
		DOC2SHSE_REIE		Identificación de emociones	0,872	
	Docente de Ciencias Sociales	DOC2SCSS_CIUACPA			Participación como ciudadano	0,748
		DOC2SCSS_CIUACPE	Prácticas que promueven la Ciudadanía activa		Percepciones sobre acciones como ciudadano	0,917
		DOC2SCSS_CIUACPE			Participación como ciudadano	0,917
DOC2SCSS_ACTEVACOM		Actividades de evaluación		Centrado en las competencias	0,856	
DOC2SCSS_ACTEVACON				Centrado en el contenido	0,813	

Grado	Actor	Código de constructo	Nombre general	Sub escala	omega
2.º grado de secundaria	Docente de Ciencias Sociales	DOC2SCSS_PROBIEE	Percepción de situaciones problemáticas en la escuela	Relacionadas al estudiante	0,943
		DOC2SCSS_PROBIER		Falta de recursos	0,814
		DOC2CSS_EDUINCG		Cognitiva	0,812
		DOC2SCSS_EDUINAF	Actitudes hacia la educación inclusiva	Afectiva	0,845
		DOC2SCSS_EDUINCD		Conductual	0,884
		DOC2SCSS_EVFORMCR	Creencias sobre la evaluación formativa	Criterios erróneos de evaluación	0,788
		DOC2SCSS_EVFORMRE		Creencias erróneas sobre retroalimentación	0,671
		DOC2SLEC_ESTRELEC	Enseñanza sobre estrategias de lectura		0,878
		DOC2SLEC_COMPDEC	Prácticas pedagógicas en Comunicación con énfasis en la repetición	Decodificación	0,701
		DOC2SLEC_COMPLECSIG		Significado	0,789
	DOC2SLEC_COMPREP	Repetición		0,878	
	DOC2SLEC_AUTLEC	Autoeficacia de docente para la enseñanza de la lectura		0,816	
	DOC2SLEC_MOTLEC	Motivación a la enseñanza de lectura		0,758	
	Docente de Comunicación	DOC2SLEC_PROBIEE	Percepción de situaciones problemáticas en la escuela	Relacionadas al estudiante	0,927
		DOC2SLEC_PROBIER		Falta de recursos	0,789
		DOC2SLEC_EDUINAF		Afectiva	0,837
		DOC2SLEC_EDUINCD	Actitudes hacia la educación inclusiva	Conductual	0,883
		DOC2SLEC_EDUINCG		Cognitiva	0,822
		DOC2SLEC_EVFORMCR	Creencias sobre la evaluación formativa	Criterios erróneos de evaluación	0,797
		DOC2SLEC_EVFORMRE		Creencias erróneas sobre retroalimentación	0,719

Grado	Actor	Código de constructo	Nombre general	Sub escala	omega	
2.º grado de secundaria	Docente de Matemática	DOC2SMAT_PROBIEE	Percepción de situaciones problemáticas en la escuela	Relacionadas al estudiante	0,893	
		DOC2SMAT_APRACT	Creencias sobre la enseñanza de la matemática	Enfoque de aprendizaje activo	0,59	
		DOC2SMAT_APRTRA		Enfoque tradicional	0,648	
		DOC2SMAT_NATPROC	Creencias sobre la naturaleza de las matemáticas	Proceso de búsqueda	0,78	
		DOC2SMAT_NATREG		Reglas y procesos	0,721	
		DOC2SMAT_PROBIER	Percepción de situaciones problemáticas en la escuela	Falta de recursos	0,759	
		DOC2SMAT_HABIN	Creencias sobre la matemática como habilidad innata		0,701	
		DOC2SMAT_DEDUINAF	Actitudes hacia la educación inclusiva	Afectiva	0,828	
		DOC2SMAT_DEDUINCG		Cognitiva	0,826	
		DOC2SMAT_EVFORMCR	Creencias sobre la evaluación formativa	Criterios erróneos de evaluación	0,733	
		DOC2SMAT_EVFORMRE		Creencias erróneas sobre retroalimentación	0,647	
		DOCS2SMAT_DEDUINCD	Actitudes hacia la educación inclusiva	Conductual	0,873	
		Estudiante	EST2SGEN_INVHOG	Involucramiento parental en el hogar		0,882
			EST2SGEN_SENPER	Sentido de pertenencia		0,802
			EST2SGEN_PERNORM	Percepción de las normas y disciplina		0,675
			EST2SGEN_INTEXT	Interés por los textos en la escuela		0,822
EST2SGEN_MOTLAU	Motivación hacia la lectura: tiempo libre		Motivación autónoma	0,897		
EST2SGEN_MOTLCO			Motivación controlada	0,821		

Grado	Actor	Código de constructo	Nombre general	Sub escala	omega
2.º grado de secundaria	Estudiante	EST2SMAT_CLIMERN	Clima del error en Matemática	Clima del error negativo	0,744
		EST2SMAT_CLIMERP		Clima del error positivo	0,614
		EST2SMAT_EVFORM	Prácticas de evaluación formativa en Matemática		0,883
		EST2SMAT_ANSIE	Ansiedad ante la matemática		0,939
		EST2SCSS_DISCU	Discusión en clases de CCSS		0,898
		EST2SCSS_CREHISTA	Pensamiento histórico del estudiante	Aplicado	0,7
		EST2SCSS_CREHISTT		Tradicional	0,626
		EST2SCSS_CIUACPA	Ciudadanía activa	Participación como ciudadano	0,872
		EST2SCSS_CIUACPE		Percepciones sobre acciones como ciudadano	0,838
		EST2SCSS_CIUACPS		Pensamiento social	0,823
		EST2SCSS_ACTEVACOM	Actividades de evaluación en CCSS	Centrado en las competencias	0,702
		EST2SCSS_ACTEVACON		Centrado en el conocimiento	0,826
		EST2SLEC_ACTCOG	Activación cognitiva en clases de Comunicación		0,901
		EST2SLEC_MOTLAU	Motivación hacia la lectura: en la escuela	Motivación autónoma	0,908
		EST2SLEC_MOTLCO		Motivación controlada	0,871
		EST2SHSE_CSEIMG	Clima social escolar	Contexto imaginativo	0,852
		EST2SHSE_CSEINS		Contexto instruccional	0,873
EST2SHSE_CSEINT	Contexto interpersonal	0,792			
EST2SHSE_CSEREG	Contexto regulatorio	0,849			

Grado	Actor	Código de constructo	Nombre general	Sub escala	omega
2.º grado de secundaria	Estudiante	EST2SHSE_EAAAE	Estrés académico adolescente	Autoexpectativa	0,766
		EST2SHSE_EAACT		Carga de trabajo	0,836
		EST2SHSE_EAADD		Desanimo/desperanza	0,763
		EST2SHSE_EAAPE		Presión por el estudio	0,723
		EST2SHSE_EAAPN		Preocupación por las notas	0,747

5.3.3. Elaboración de puntajes

Una vez confirmada la consistencia interna de los indicadores recogidos, se procedió a generar los puntajes para cada uno de los actores educativos. Los puntajes permiten representar el nivel de presencia de la variable latente en cada uno de los encuestados.

Una de las formas más simples de establecer estos puntajes es mediante la suma o promedio simple con todos los ítems que conforman el constructo. Sin embargo, todos los ítems que conforman un factor no aportan lo mismo para la formación del constructo, pues tienen diferentes pesos o cargas factoriales. En consecuencia, de elaborarse puntajes a través de suma o promedio simples, se le otorgaría el mismo peso a todos los ítems, lo que haría que ítems con valores de carga relativamente bajos reciban igual importancia que los de carga alta para el cálculo de los puntajes. En consecuencia, ignorar las diferentes cantidades de variabilidad en las variables observadas puede dar lugar a puntuaciones factoriales menos fiables (DiStefano et al., 2009).

Debido a ello, para el caso de los constructos reflexivos, los puntajes fueron estimados mediante el método EBM (Empirical Bayes Modal), el cual es adecuado para ítems categóricos. Este procedimiento se aplicó con la función *lavPredict* del paquete *lavaan* del software estadístico R. Para los constructos formativos, los ítems fueron estandarizados y luego multiplicados por el peso asociado al primer componente. Una vez estimados los puntajes, estos fueron estandarizados con una media = 0 y desviación estándar = 1. La transformación de los puntajes de la variable permite una interpretación más sencilla cuando son incluidos en los modelos de regresión.

Referencias

Referencias

- American Educational Research Association, American Psychological Association y National Council of Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573. <https://doi.org/10.1007/BF02293814>
- Babbie, E. (2010). *The practice of social research*. Wadsworth Cengage Learning.
- Baker F. B. y Kim S. H. (2017). *The basics of item response theory using R*. Springer International Publishing
- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.
- Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, 16(3), 265–284. <https://doi.org/10.1037/a0024448>
- Bond, T. G., Yan, Z. y Heene, M. (2021). *Applying the Rasch model. Fundamental measurement in the human Sciences (4a ed.)*. Routledge
- Browning, D. (2003). Common definitions: Adaptations, accomodations, modifications. http://www.pent.ca.gov/acc/commondefinitions_accom-mod.pdf
- Charter, R. A. (2003). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *The Journal of General Psychology*, 130(3), 290-304. <https://doi.org/10.1080/00221300309601160>
- Chen, F., Carolina, N., Curran, P. J., Bollen, K. A., & Kirby, J. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Social Methods Research*, 36(4), 1-31. <https://doi.org/10.1177/0049124108314720>
- Cizek, G. J. (1993). Reconsidering standards and criteria. *Journal of Educational Measurement*, 30(2), 93-106. <https://doi.org/10.1111/j.1745-3984.1993.tb01068.x>
- Cizek, G. J., y Bunch, M. B. (2007). *Standard settings*. SAGE.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. <https://doi.org/10.1007/BF02310555>
- De Ayala, R. (2022). *The theory and practice of Item Response Theory (2nd ed.)*. The Guilford Press.
- DeMars, C. (2010). *Item response theory. Understanding statistics measurement*. Oxford University Press, Inc.
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.

- DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 425–438. <https://doi.org/10.1080/10705511.2014.915373>
- DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation*, 14, Article 20. <https://doi.org/10.7275/da8t-4g52>
- Du Toit M. (2003). IRT from SSI: BILOG-MG MULTILOG PARSCALE TESTFACT. Scientific Software International, Inc.
- Embretson, S., y Reise, S. (2000). *Item Response Theory for psychologists*. Lawrence Erlbaum Associates.
- Fisher, W. (1992). Reliability statistics. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 2*, 1996 (p. 238). MESA Press.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466–491. <https://doi.org/10.1037/1082-989X.9.4.466>
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39(2), 291-314.
- Fox, C. (1999). An introduction to the Partial Credit model for developing nursing assessments. *Journal of Nursing Education*, 38(8), 340-346. <https://doi.org/10.3928/0148-4834-19991101-04>
- Furr, R. M. (2013). *Scale construction and psychometrics for social and personality psychology*. SAGE Publications.
- Hambleton, R., Swaminathan, H. y Rogers, J. (1991). *Fundamentals of Item Response Theory*. SAGE.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139-164. <https://doi.org/10.1177/014662168500900204>
- Ho, C., y Osborn, S. E. (2005). Test equating by common items and common subjects: Concepts and applications. *Practical Assessment, Research & Evaluation*, 10, Article 4. <https://doi.org/10.7275/68dy-z131>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Ingebo, G. (1997). *Probability in the measure of achievement*. MESA Press.
- Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30(2), 199-218. <https://doi.org/10.1086/376806>

- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, 374, 1-16. <https://doi.org/10.1098/rsta.2015.0202>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling*. R package version 0.5-6. <https://CRAN.R-project.org/package=semTools>
- Karantonis, A., y Sireci, S. G. (2006). The bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice*, 25(1), 4-12. <https://doi.org/10.1111/j.1745-3992.2006.00047.x>
- Kim, D., Choi, S., Um, K. y Kim, J. (2006, Abril). A comparison of methods for estimating classification consistency. Trabajo presentado en la Annual Meeting of the National Council on Education in Measurement, San Francisco, CA.
- Kish, L. (1972). *Muestreo de encuestas*. Trillas.
- Kolen, M., y Brennan, R. (2014). *Test equating, scaling and linking. Methods and practices* (3a ed.). Springer.
- Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research*, 51(2-3), 220-239. <https://doi.org/10.1080/00273171.2015.1134306>
- Lee, W. (2010). Classification consistency and accuracy for complex assessments using Item Response Theory. *Journal of Educational Measurement*, 47(1), 1-17. <https://doi.org/10.1111/j.1745-3984.2009.00096.x>
- Leppink, J. (2019). *Statistical methods for experimental research in education and psychology*. Springer.
- Lewis, D. M., Mitzel, H. C. y Green, D. R. (1996, Junio). Standard setting: A bookmark approach. En D. R. Green (Ed.), *IRT-based standard-setting procedures utilizing behavioral anchoring*. Simposio presentado en el Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 5-55. https://legacy.voteview.com/pdf/Likert_1932.pdf
- Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, 2(3), 266-283.
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3(2), 103-122.
- Linacre, J. M. (2001). Category, step and threshold: Definitions & disordering. *Rasch Measurement Transactions*, 15(1), 794. <https://www.rasch.org/rmt/rmt151g.htm>
- Linacre, J. M. (2024). A user's guide to WINSTEPS, MINISTEP Rasch model computer programs. <https://www.winsteps.com/a/Winsteps-Manual.pdf>

- Livingston, S. A. (2004). Equating test scores (without IRT). Educational Testing Service. <https://www.ets.org/Media/Research/pdf/LIVINGSTON2ed.pdf>
- Lord, F. M. & Novick, M. R. (1968). Statistical theories of mental tests scores. Addison-Wesley.
- Lohr, S. (2021). *Samplig. Design and analysis*. CRC Press.
- Lumley, T. (2010) *Complex surveys: A guide to analysis using R*. John Wiley and Sons.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- Mair, P. (2018). *Modern psychometrics with R*. Springer. https://doi.org/10.1007/978-3-319-93177-7_6
- Malhotra, N. (2006). Questionnaire design and scale development. En: R. Grover, & M. Vriens (Eds.), *The handbook of marketing research: Uses, misuses, and future advances* (pp. 84-94). SAGE Publications. <https://doi.org/10.4135/9781412973380>
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174. <https://doi.org/10.1007/BF02296272>
- McGartland D., Berg-Weger, M., Tebb, S. S., Lee, E. S. & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104. <https://doi.org/10.1093/swr/27.2.94>
- Ministerio de Educación. (2005). Evaluación nacional del rendimiento estudiantil 2004. Informe descriptivo de resultados. <http://repositorio.minedu.gob.pe/handle/123456789/1134>
- Ministerio de Educación. (2009). Evaluación Censal de Estudiantes (ECE). Segundo grado de primaria y cuarto grado de primaria de IE EIB. Marco de trabajo. <http://repositorio.minedu.gob.pe/handle/123456789/1106>
- Ministerio de Educación. (2015). Marco de fundamentación de las pruebas de rendimiento de la Evaluación Censal de Estudiantes de 2do grado de secundaria 2015. <http://repositorio.minedu.gob.pe/handle/123456789/5055>
- Moreno, R., Martínez, R. J. y Muñiz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema*, 16(3), 490–497. <https://www.psicothema.com/pi?pii=3023>
- Muñiz, J. (1998). *Teoría clásica de los tests* (2nd ed.). Pirámide.

- Muthén, B., du Toit, S., & Spisic, D. (1997). Robust interference using weighted least squared and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished Tech. Rep. No. Los Angeles, CA.
- Navas, M. (1996). Equiparación de puntuaciones. En J. Muñiz (Ed.), *Psicometría*, (pp. 293-370). Universitas.
- Nunnally, J., y Bernstein, I. (1995). *Teoría psicométrica* (3rd ed.). McGraw-Hill.
- OECD (2024). PISA 2022 Technical Report. OECD Publishing. <https://doi.org/10.1787/01820d6d-en>.
- Orlando M. y Thissen D. (2000). Likelihood-based item-fit indices for dichotomous Item Response Theory models. *Applied Psychological Measurement*, 24(1), 50-64. <https://journals.sagepub.com/doi/10.1177/01466216000241003>
- Park, H.S., Dailey, R. & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, 28(4), 562-577. <https://doi.org/10.1111/j.1468-2958.2002.tb00824.x>
- Patel, H. R., & Joseph, J. M. (2016). Questionnaire designing process: A review. *Journal of Clinical Trials*, 6(2), 2-7. <https://doi.org/10.4172/2167-0870.1000255>
- Paz, M. (1996). Validez. En: J. Muñiz (Ed.), *Psicometría* (pp. 49-103). Universitas.
- Pérez, L. (2006). *Muestreo estadístico: Conceptos y problemas resueltos*. Pearson Education.
- Preacher, K.J. & McCallum, R.C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2(1), 13-32. https://doi.org/10.1207/S15328031US0201_02
- Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Nielsen & Lydiche.
- Ravela, P. (2006). *Fichas didácticas para comprender las evaluaciones educativas*. Editorial San Marino. <https://hdl.handle.net/20.500.12799/4999>
- Revelle, W. (2024). How to use the psych package for regression and mediation analysis. <https://cloud.r-project.org/web/packages/psychTools/vignettes/mediation.pdf>
- Robinson, M. A. (2018). Using multi-item psychometric scales for research and practice in human resource management. *Human Resource Management*, 57(3), 739-750. <https://doi.org/10.1002/hrm.21852>
- Rosseel, Y. (2012). lavaan: An R package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Rutkowski, L., & Svetina, D. (2013). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31-57. <https://doi.org/10.1177/0013164413498257>

- Schulz, E. M. (1990). Functional assessment of fit. En J. M. Linacre (Ed.), *Rasch Measurement Transactions*, 1990 (p. 82). MESA Press
- Schumacker, R. E., y Smith, E. V. (2007). Reliability: A Rasch perspective. *Educational and Psychological Measurement*, 67(3), 394-409. <https://doi.org/10.1177/0013164406294776>
- Sheaffer, R. L., Mendelhall, W., & Ott, L. (2006). *Elementary survey sampling* (6th ed.). Thomson Brooks
- Shepard, L. (1980). Standard setting issues and methods. *Applied Psychological Measurement*, 4(4), 447-467. <https://doi.org/10.1177/014662168000400403>
- Smith, R., y Kramer, G. (1989). Response pattern analysis with supplemental score reports. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1*, 1995 (pp.33-35). MESA Press.
- Stehr-Green, P., Stehr-Green, J., & Nelson, A. (2003). Developing a questionnaire. *Focus on Field Epidemiology*, 2(2), 1-6. https://nciph.sph.unc.edu/focus/vol2/issue2/2-2Questionnaire_issue.pdf
- Suen, H. (1990). *The psychometric process*. Pennsylvania State University.
- Willis, G. B. (2008). Cognitive interviewing. En: *Encyclopedia of Survey Research Methods* (pp. 107-109). Sage Publications. <https://doi.org/10.4135/9781412963947>
- Wilson, M. (2005). *Constructing measures. An item response modeling approach*. Lawrence Erlbaum Associates.
- Wind, S., y Hua, C. (2022). Rasch measurement theory analysis in R. Chapman & Hall.
- Wright, B. & Linacre, J. (1989). Observations are always ordinal; Measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, 70(12), 857-860.
- Wright, B. D. (1999). Model selection: Rating Scale Model (RSM) or Partial Credit Model (PCM)? *Rasch Measurement Transactions*, 12(3), 641-642. <https://www.rasch.org/rmt/rmt1231.htm>
- Wright, B. D., y Linacre, J. M. (1994). Reasonable mean-square fit values. En J. M. Linacre (Ed.), *Rasch Measurement Transactions*, 1994 (p. 370). MESA Press
- Wright, B. D., y Stone, M. (1998). *Diseño de mejores pruebas*. CENEVAL.
- Wright, B. D., y Masters, G. (1982). *Rating scale analysis*. MESA Press.
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2-3), 114-128. <https://doi.org/10.1016/j.stueduc.2005.05.005>
- Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise and Sport*, 69(1), 11-23. <https://doi.org/10.1080/02701367.1998.10607662>

Anexos

Anexo A Especificaciones ENLA 2023. Lectura – 2.º grado de primaria
Tabla A.1 Distribución de ítems por capacidad en la prueba

Capacidad	Total	Porcentaje de la prueba
Se apropia del sistema de escritura (Lectura inicial).	6	6,7 %
Obtiene información del texto escrito.	31	34,4 %
Infiere e interpreta información del texto.	39	43,3 %
Reflexiona y evalúa la forma, el contenido y el contexto del texto.	14	15,6 %
TOTAL	90	100 %

Tabla A.2 Distribución de textos en la prueba

Tipo de texto	Género	Correlativo del texto	Cantidad por tipo de texto	Porcentaje por tipo de texto		
Narrativo	Cuento	1	4	15,4 %		
	Leyenda	2				
	Anécdota	3				
		4				
Instructivo	Manual de procedimientos	5	2	7,7 %		
	Texto de recomendaciones	6				
Descriptivo		7	6	23,1 %		
	Artículo enciclopédico				8	
					9	
					10	
					11	
	Cuadro comparativo	12				
Argumentativo	Afiche	13	4	15,4 %		
		14				
	15					
	Artículo de opinión	16				
No tipificable Palabra y oración	Nota	17	10	38,5 %		
	Aviso	18				
		19				
	Palabra y oración	20			8	30,8 %
		21				
		22				
					23	

No tipificable Palabra y oración	Palabra y oración	24	10	38,5 %
		25		
		26		
TOTAL			26	100 %

Tabla A.3 Distribución de ítems en la prueba

Tipo de texto	Género	Correlativo del texto	Cantidad de ítems por cada texto	Cantidad de ítems por tipo de texto	Porcentaje de ítems por tipo de texto
Narrativo	Cuento	1	5	18	20,0 %
	Leyenda	2	5		
	Anécdota	3	4		
		4	4		
Instructivo	Manual de procedimientos	5	6	11	12,2 %
	Texto de recomendaciones	6	5		
Descriptivo	Artículo enciclopédico	7	4	28	31,1 %
		8	6		
		9	6		
		10	3		
		11	5		
	Cuadro comparativo	12	4		
Argumentativo	Afiche	13	4	18	20,0 %
		14	6		
		15	4		
	Artículo de opinión	16	4		
No tipificable Palabra y oración	Nota	17	4	15	16,7 %
	Aviso	18	3		
	Palabra y oración	19	1		
		20	1		
		21	1		
		22	1		
		23	1		
		24	1		
25	1				
26	1				
TOTAL			90	90	100,0 %

Anexo B Especificaciones EM 2023. Lectura – 4.° grado de primaria
Tabla B.1 Distribución de ítems por capacidad en la prueba

Capacidad	Total	Porcentaje de la prueba
Obtiene información del texto escrito.	28	29 %
Infiere e interpreta información del texto.	44	46 %
Reflexiona y evalúa la forma, el contenido y el contexto del texto.	24	25 %
TOTAL	96	100 %

Tabla B.2 Distribución de textos en la prueba

Tipo de texto	Género	Correlativo del texto	Cantidad por tipo de texto	Porcentaje por tipo de texto	
Narrativo	Cuento	1	5	29 %	
	Historieta	2			
	Noticia	3			
	Anécdota	4			
		5			
Descriptivo	Artículo enciclopédico	6	5	29 %	
		7			
		8			
		Artículo periodístico			9
		Infografía			10
Instructivo	Manual de procedimientos	11	2	12 %	
		12			
Argumentativo	Infografía	13	3	18 %	
	Afiche	14			
	Artículo de opinión	15			
Expositivo	Artículo de divulgación	16	2	12 %	
		17			
TOTAL			17	100 %	

Tabla B.3 Distribución de ítems en la prueba

Tipo de texto	Género	Correlativo del texto	Cantidad de ítems por cada texto	Cantidad de ítems por tipo de texto	Porcentaje de ítems por tipo de texto
Narrativo	Cuento	1	6	24	25,0 %
	Historieta	2	6		
	Noticia	3	6		
	Anécdota	4	3		
		5	3		
Descriptivo	Artículo enciclopédico	6	6	30	31,0 %
		7	6		
		8	6		
	Artículo periodístico	9	6		
	Infografía	10	6		
Instructivo	Manual de procedimientos	11	6	12	12,5 %
		12	6		
Argumentativo	Infografía	13	6	18	19,0 %
	Afiche	14	6		
	Artículo de opinión	15	6		
Expositivo	Artículo de divulgación	16	6	12	12,5 %
		17	6		
TOTAL				96	100 %

Anexo C Especificaciones EM 2023. Lectura – 2.º grado de secundaria
Tabla C.1 Distribución de ítems por capacidad en la prueba

Capacidad	Total	Porcentaje de la prueba
Obtiene información del texto escrito.	25	20,5 %
Infiere e interpreta información del texto.	58	47,5 %
Reflexiona y evalúa la forma, el contenido y el contexto del texto.	39	32,0 %
TOTAL	122	100 %

Tabla C.2 Distribución de textos en la prueba

Tipo de texto	Género	Correlativo del texto	Cantidad de ítems por cada texto	Porcentaje por tipo de texto
Narrativo	Crónica	1	3	15,0 %
		2		
	Historieta	3		
Instructivo	Manual de procedimientos	4	3	15,0 %
		5		
		6		
Descriptivo	Artículo enciclopédico	7	3	15,0 %
		8		
		9		
Argumentativo	Artículo de opinión	10	6	30,0 %
		11		
		12		
		13		
		14		
		15		
Expositivo	Artículo de divulgación	16	5	25,0 %
		17		
		18		
		19		
	Afiche	20		
TOTAL			20	100 %

Tabla C.3 Distribución de textos en la prueba

Tipo de texto	Género	Correlativo del texto	Cantidad de ítems por cada texto	Cantidad de ítems por tipo de texto	Porcentaje por tipo de texto
Narrativo	Crónica	1	6	18	14,8 %
		2	6		
	Historieta	3	6		
Instructivo	Manual de procedimientos	4	6	18	14,8 %
		5	6		
	Texto de recomendaciones	6	6		
Descriptivo	Artículo enciclopédico	7	7	19	15,6 %
		8	6		
	Infografía	9	6		
Argumentativo	Artículo de opinión	10	6	36	29,5 %
		11	6		
		12	6		
		13	6		
	Reseña	14	6		
	Infografía	15	6		
Expositivo	Artículo de divulgación	16	7	31	25,4 %
		17	6		
		18	6		
		19	6		
	Afiche	20	6		
TOTAL			122	122	100 %

Anexo D Especificaciones ENLA 2023. Matemática – 2.º grado de primaria
Tabla D.1 Matemática ENLA 2023. Distribución de ítems por competencias en la prueba

Competencias	Cantidad de ítems	Porcentaje
Resuelve problemas de cantidad.	92	100 %
Resuelve problemas de regularidad, equivalencia y cambio.	0	0 %
Resuelve problemas de forma, movimiento y localización.	0	0 %
Resuelve problemas de gestión de datos e incertidumbre.	0	0 %
Total	92	100 %

Tabla D.2 Matemática ENLA 2023. Distribución de ítems por capacidades en la prueba

Competencia	Capacidades	Cantidad de ítems	Porcentaje
Resuelve problemas de cantidad	Traduce cantidades a expresiones numéricas.	46	50,0 %
	Comunica su comprensión sobre los números y las operaciones.	19	20,7 %
	Usa estrategias y procedimientos de estimación y cálculo.	15	16,3 %
	Argumenta afirmaciones sobre las relaciones numéricas y las operaciones.	12	13,0 %
Total		92	100 %

El total ha sido redondeado a un decimal, puesto que la suma en esta tabla no da exactamente 100 %.

Tabla D.3 Matemática ENLA 2023. Distribución de ítems por contextos en la prueba

Contextos	Cantidad de ítems	Porcentaje
Intramatemático	9	9,8 %
Extramatemático	83	90,2 %
Total	92	100 %

Tabla E.1 Matemática ENLA 2023. Distribución de ítems por competencias en la prueba

Competencias	Cantidad de ítems	Porcentaje
Resuelve problemas de cantidad.	39	38,2 %
Resuelve problemas de regularidad, equivalencia y cambio.	26	25,5 %
Resuelve problemas de forma, movimiento y localización.	20	19,6 %
Resuelve problemas de gestión de datos e incertidumbre.	17	16,7 %
Total	102	100 %

Tabla E.2 Matemática ENLA 2023. Distribución de ítems por capacidades en la prueba

Competencia	Capacidades	Cantidad de ítems	Porcentaje
Resuelve problemas de cantidad.	Traduce cantidades a expresiones numéricas.	20	19,6 %
	Comunica su comprensión sobre los números y las operaciones.	15	14,7 %
	Usa estrategias y procedimientos de estimación y cálculo.	3	2,9 %
	Argumenta afirmaciones sobre las relaciones numéricas y las operaciones.	1	1,0 %
Resuelve problemas de regularidad, equivalencia y cambio.	Traduce datos y condiciones a expresiones algebraicas y gráficas.	3	2,9 %
	Comunica su comprensión sobre las relaciones algebraicas.	3	2,9 %
	Usa estrategias y procedimientos para encontrar equivalencias y reglas generales.	14	13,7 %
Resuelve problemas de forma, movimiento y localización.	Argumenta afirmaciones sobre las relaciones de cambio y equivalencia.	6	5,9 %
	Modela objetos con formas geométricas y sus transformaciones.	4	3,9 %
	Comunica su comprensión sobre las formas y relaciones geométricas.	10	9,8 %
	Usa estrategias y procedimientos para orientarse en el espacio.	3	2,9 %
	Argumenta afirmaciones sobre relaciones geométricas.	3	2,9 %

Competencia	Capacidades	Cantidad de ítems	Porcentaje
Resuelve problemas de gestión de datos e incertidumbre.	Representa datos con gráficos y medidas estadísticas o probabilísticas.	1	1,0 %
	Comunica su comprensión de los conceptos estadísticos y probabilísticos.	8	7,8 %
	Usa estrategias y procedimientos para recopilar y procesar datos.	3	2,9 %
	Sustenta conclusiones o decisiones con base en la información obtenida.	5	4,9 %
Total		102	100,0 %

El total ha sido redondeado a un decimal, puesto que la suma en esta tabla no da exactamente 100 %.

Tabla E.3 Matemática ENLA 2023. Distribución de ítems por contextos en la prueba

Contextos	Cantidad de ítems	Porcentaje
Intramatemático	14	13,7 %
Extramatemático	88	86,3 %
Total	102	100,0 %

Tabla F.1 Matemática ENLA 2023. Distribución de ítems por competencias en la prueba

Competencias	Cantidad de ítems	Porcentaje
Resuelve problemas de cantidad.	33	27,0 %
Resuelve problemas de regularidad, equivalencia y cambio.	32	26,2 %
Resuelve problemas de forma, movimiento y localización.	29	23,8 %
Resuelve problemas de gestión de datos e incertidumbre.	28	23,0 %
Total	122	100,0 %

Tabla F.2 Matemática ENLA 2023. Distribución de ítems por capacidades en la prueba

Competencia	Capacidades	Cantidad de ítems	Porcentaje
Resuelve problemas de cantidad.	Traduce cantidades a expresiones numéricas.	17	13,9 %
	Comunica su comprensión sobre los números y las operaciones.	9	7,4 %
	Usa estrategias y procedimientos de estimación y cálculo.	4	3,3 %
	Argumenta afirmaciones sobre las relaciones numéricas y las operaciones.	3	2,5 %
Resuelve problemas de regularidad, equivalencia y cambio.	Traduce datos y condiciones a expresiones algebraicas y gráficas.	14	11,5 %
	Comunica su comprensión sobre las relaciones algebraicas.	4	3,3 %
	Usa estrategias y procedimientos para encontrar equivalencias y reglas generales.	9	7,4 %
Resuelve problemas de forma, movimiento y localización.	Argumenta afirmaciones sobre relaciones de cambio y equivalencias.	5	4,1 %
	Modela objetos con formas geométricas y sus transformaciones.	7	5,7 %
	Comunica su comprensión sobre las formas y relaciones geométricas.	8	6,6 %
	Usa estrategias y procedimientos para orientarse en el espacio.	10	8,2 %
	Argumenta afirmaciones sobre relaciones geométricas.	4	3,3 %

Competencia	Capacidades	Cantidad de ítems	Porcentaje
Resuelve problemas de gestión de datos e incertidumbre.	Representa datos con gráficos y medidas estadísticas o probabilísticas.	4	3,3 %
	Comunica su comprensión de los conceptos estadísticos y probabilísticos.	12	9,8 %
	Usa estrategias y procedimientos para recopilar y procesar datos.	5	4,1 %
	Sustenta conclusiones o decisiones con base en la información obtenida.	7	5,7 %
Total		122	100,0 %

El total ha sido redondeado a un decimal, puesto que la suma en esta tabla no da exactamente 100 %.

Tabla F.3 Matemática ENLA 2023. Distribución de ítems por contextos en la prueba

Contextos	Cantidad de ítems	Porcentaje
Intramatemático	6	4,9 %
Extramatemático	116	95,1 %
Total	122	100,0 %

Tabla G.1 Ciencias Sociales ENLA 2023. Distribución de ítems por competencias en la prueba

Competencia	Cantidad de ítems por competencia	Porcentaje de ítems
Construye interpretaciones históricas.	49	38,9 %
Gestiona responsablemente el espacio y el ambiente.	45	35,7 %
Gestiona responsablemente los recursos económicos.	32	25,4 %
Total	126	100,0 %

Tabla G.2 Ciencias Sociales ENLA 2023. Distribución de ítems por capacidad en la prueba

Competencia	Capacidad	Cantidad de ítems por capacidad	Porcentaje de ítems por capacidad
Construye interpretaciones históricas.	Interpreta críticamente fuentes diversas.	14	28,6 %
	Comprende el tiempo histórico.	15	30,6 %
	Elabora explicaciones sobre procesos históricos.	20	40,8 %
Total		49	100,0 %
Gestiona responsablemente el espacio y el ambiente.	Maneja fuentes de información para comprender el espacio geográfico y el ambiente.	9	20,0%
	Comprende las relaciones entre los elementos naturales y sociales.	24	53,3 %
	Genera acciones para conservar el ambiente local y global.	12	26,7 %
Total		45	100,0 %
Gestiona responsablemente los recursos económicos.	Comprende el funcionamiento del sistema económico y financiero.	11	34,4 %
	Toma decisiones económicas y financieras.	21	65,6 %
Total		32	100,0 %

El total ha sido redondeado a un decimal, puesto que la suma en esta tabla no da exactamente 100 %.

Tabla G.3 Ciencias Sociales ENLA 2023. Distribución de ítems por conocimientos en la prueba

Competencia	Conocimiento	Cantidad de ítems	Porcentaje de ítems por conocimiento
Construye interpretaciones históricas.	Evolución y mundo antiguo	5	10,2 %
	Poblamiento y desarrollo en los Andes	9	18,4 %
	Edad media y feudalismo	11	22,4 %
	Edad moderna y encuentro de dos mundos	9	18,4 %
	Conquista e inicios del virreinato	11	22,4 %
	Nociones básicas de la historia	4	8,2 %
	Total	49	100,0 %
Gestiona responsablemente el espacio y el ambiente.	Nociones generales de geografía	7	15,6 %
	Espacio geográfico y territorio	9	20,0 %
	Desarrollo sostenible	15	33,3 %
	Gestión de riesgo de desastre	9	20,0 %
	Problemáticas ambientales	5	11,1 %
	Total	45	100 %
Gestiona responsablemente los recursos económicos.	Formalidad económica	5	15,6 %
	Funcionamiento del mercado	8	25,0 %
	Gestión responsable de las finanzas	10	31,3 %
	Deberes y derechos de los agentes económicos	5	15,6 %
	Rol de los agentes económicos	4	12,5 %
Total	32	100,0 %	

El total ha sido redondeado a un decimal, puesto que la suma en esta tabla no da exactamente 100 %.

Tabla G.4 Ciencias Sociales ENLA 2023. Distribución de ítems por contexto en la prueba

Contexto	Cantidad de ítems	Porcentaje de ítems por contexto
Personal/Familiar	21	16,7 %
Escolar/Local	29	23,0 %
Nacional	43	34,1 %
Internacional	33	26,2 %
Total	126	100,0 %

Anexo H Especificaciones ENLA 2023. Lectura en castellano como segunda lengua – 4.º grado de primaria EIB

Tabla H.1 Distribución de ítems por capacidad en la prueba

Capacidad	Total	Porcentaje de la prueba
Se apropia del sistema de escritura (Lectura inicial).	48	33,3 %
Obtiene información del texto escrito.	38	26,4 %
Infiere e interpreta información del texto.	43	29,9 %
Reflexiona y evalúa la forma, el contenido y el contexto del texto.	15	10,4 %
TOTAL	144	100 %

Tabla H.2 Distribución de textos en la prueba

Tipo de texto	Género	Correlativo del texto	Cantidad por tipo de texto	Porcentaje por tipo de texto	
Narrativo	Cuento	1	9	10,8 %	
		2			
		Leyenda			3
		Noticia			4
	Anécdota	5			
		6			
		7			
		8			
		9			
Instructivo	Manual de procedimientos	10	2	2,4 %	
		11			
Descriptivo	Artículo enciclopédico	12	10	12,0 %	
		13			
		14			
		15			
		16			
		17			
		18			
		19			
		20			
		21			

Tipo de texto	Género	Correlativo del texto	Cantidad por tipo de texto	Porcentaje por tipo de texto
Argumentativo	Artículo de opinión	22	1	1,2 %
Expositivo	Artículo de divulgación	23	1	1,2 %
No tipificable	Palabra y oración	24 - 83	60	72,3 %
TOTAL			83	100 %

Tabla H.3 Distribución de ítems en la prueba

Tipo de texto	Género	Correlativo del texto	Cantidad de ítems por cada texto	Cantidad de ítems por texto	Porcentaje por de ítems por tipo de texto
Narrativo	Cuento	1	6	36	25,0 %
		2	5		
		3	5		
	Anécdota	4	5		
		5	3		
		6	3		
		7	3		
		8	3		
		9	3		
Instructivo	Manual de procedimientos	10	6	12	8,3 %
		11	6		
Descriptivo	Cuadro comparativo	12	3	25	17,4 %
		13	1		
		14	1		
		15	1		
		16	1		
		17	1		
		18	1		
		19	6		
		20	5		
Argumentativo	Artículo de opinión	21	5	6	4,2 %
		22	6		
		23	5		
		24 - 83	60 ¹		
		60	60		
Expositivo	Artículo de divulgación	23	5	5	3,5 %
No tipificable	Palabra y oración	24 - 83	60 ¹	60	41,7 %
TOTAL			83	144	100 %

¹ Para constructos con varias dimensiones el mínimo número que se acepta para formar una dimensión es tres debido a que en este tipo de modelo se pueden obtener soluciones que converjan.

Ministerio de Educación

**Calle Del Comercio 193,
San Borja - Lima, Perú
Tel.: (511) 615-5800**

<http://www.minedu.gob.pe>
