

COMPETENCIA
LECTURA
MATEMÁTICA
CIENCIA
SOCIOEMOCIONAL
MEDICIÓN
DESEMPEÑO
PSICOMETRÍA
APRENDIZAJE
EDUCACIÓN
CALIDAD

Reporte técnico de la Evaluación Nacional de Logros de Aprendizaje de Estudiantes (ENLA) 2024



PERÚ

Ministerio
de Educación



COMPETENCIA
LECTURA
MATEMÁTICA
CIENCIA
SOCIOEMOCIONAL
MEDICIÓN
DESEMPEÑO
PSICOMETRÍA
APRENDIZAJE
EDUCACIÓN
CALIDAD

Reporte técnico de la Evaluación
Nacional de Logros de Aprendizaje
de Estudiantes (ENLA) 2024



PERÚ

Ministerio
de Educación





PERÚ

Ministerio
de Educación

Morgan Niccolo Quero Gaime

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

Andrea Nieto García

Jefa de la Oficina de Medición de la Calidad de los Aprendizajes

**Reporte técnico de la Evaluación Nacional de Logros de Aprendizaje
(ENLA) 2024**

Responsables del documento

Andrés Alberto Burga León (Coordinador)

Yoni Arámbulo Mogollón

Olimpia Castro Mora

Adolfo Zárate Pérez

Luis Mejía Campos

Gabriela Santibáñez Rojas

Sandra Elizabeth Flores Ari

Marco Aurelio Carranza Rodríguez

Christian Alexis Moreau Almaraz

Victor Wilfredo Salazar Oré

Diego Orlando Hermoza Cárdenas

María Teresa Estefanía Sánchez

Carlos Ivan Flores Venturi

Natalia Mónica Gómez Alvarado

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, 2025

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	6
Capítulo 1: Diseño y construcción de las pruebas aplicadas	9
1.1 Construcción de los ítems y las pruebas	10
1.2 Procesos de validación de los instrumentos de evaluación	13
1.2.1 Validación de expertos de la UMC	14
1.2.2 Validación de campo: aplicación piloto.....	15
1.2.3 Análisis de las propiedades psicométricas y de la distribución de los ítems en la escala.....	15
1.2.4 Validación de expertos de otras oficinas del Minedu.....	16
1.2.5 Adaptación de instrumentos para estudiantes con discapacidad	17
1.2.6 Aplicación definitiva.....	20
1.3 Ensamblaje de las pruebas mediante el diseño bloques	20
1.3.1 Diseño de bloques de la ENLA 2024	21
Capítulo 2: Población a evaluar	33
2.1 Población objetivo	33
2.2 Marco censal y marco muestral	33
2.3 Diseño muestral	34
2.4 Cobertura en la ENLA 2024.....	37
2.5 Pesos muestrales ENLA 2024	39
2.6 Muestra de 2.º de secundaria	42
Capítulo 3: Operativo de campo	45
3.1 Organización territorial	46
3.2 Conformación de la Red Administrativa de campo	48
3.3 Proceso de selección y capacitación del personal de la RA.....	51
3.4 Características de los instrumentos y documentos utilizados	56
3.5 Proceso de aplicación	58
Capítulo 4: Procesamiento de datos y análisis psicométrico	63
4.1 Gestión y depuración de datos	63
4.1.1 Gestión física	63
4.1.2 Digitalización y captura de datos	64
4.1.3 Depuración de datos	65

4.2 Estrategias de análisis psicométrico.....	66
4.2.1 Modelo Rasch para ítems dicotómicos	66
4.2.2 Modelo Rasch de crédito parcial	69
4.2.3 Análisis de ítems.....	73
4.2.4 Evidencias de validez vinculadas a la estructura interna	74
4.2.5 Confiabilidad y consistencia de la clasificación.....	76
4.3 Resultados del análisis psicométrico en Lectura, Matemática	76
4.3.1 Calibración de los ítems	76
4.3.2 Evidencias de validez vinculadas a la estructura interna	89
4.3.3 Confiabilidad y consistencia de la clasificación.....	90
4.4 Equiparación de medidas	91
4.5 Niveles de logro y preparación de resultados.....	97
 Capítulo 5: Procesamiento de datos y elaboración de cuestionarios de	
Factores Asociados	101
5.1 Elaboración de las preguntas del cuestionario de factores asociados	101
5.2 Elaboración de análisis descriptivo.....	105
5.3 Validez y confiabilidad de los índices y escalas de factores asociados	105
5.3.1 Evidencias de validez basadas en la estructura interna.....	106
5.3.2 Evidencias de Confiabilidad.....	119
5.4 Elaboración de puntajes.....	130
 Capítulo 6: Diseño, construcción y procesamiento de datos de Habilidades	
Socioemocionales	132
6.1 Conceptos clave.....	133
6.2 Procedimiento para la elaboración del instrumento.....	134
6.3 Análisis psicométricos para habilidades socioemocionales.....	134
6.3.1 Calibración final de los ítems.....	135
6.3.2 Estructura de umbrales	136
6.3.3 Evidencias de validez vinculadas a la estructura interna	137
6.3.4 Confiabilidad.....	137
6.4 Presentación de resultados	138
 Referencias	140
 Anexos	148

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 120-2024-Minedu que aprueba la Evaluación Nacional de Logros de Aprendizaje – ENLA 2024, con el objetivo de obtener un diagnóstico actualizado de los aprendizajes de los estudiantes.

Esta evaluación se aplicó de manera presencial a estudiantes de 4.º grado de primaria (carácter censal) y 6.º grado de primaria (carácter muestral) de instituciones educativas públicas y privadas de todo el país.

En la ENLA 2024 se evaluó en 4.º y 6.º grados de primaria las siguientes competencias: Comunicación (lee diversos tipos de textos escritos en su lengua materna) y Matemática (todas las competencias). Además, y como parte de un enfoque integral para comprender y fortalecer el bienestar de los estudiantes en el país, en estudiantes de 6.º grado de primaria se evaluó la habilidad socioemocional de Autorregulación, entendida como la articulación entre conciencia emocional, regulación emocional y regulación conductual.

La ENLA 2024 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.

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 la evaluación del 2024.

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.

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.

Finalmente, en el sexto capítulo se menciona el proceso del diseño, la construcción y el procesamiento de datos de la medición en habilidades socioemocionales, específicamente de la habilidad de Autorregulación, considerada en la Evaluación Nacional de Logros de Aprendizaje (ENLA) 2024.

Diseño y construcción de las pruebas aplicadas

Capítulo 1

Capítulo 1

Diseño y construcción de las pruebas aplicadas

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. 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 2024, mediante R.M. 120-2024-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 muestral y censal implementadas por la UMC en el 2024

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.	

Nivel	Grado	Área	Competencia(s) evaluada(s)	Denominación de la prueba
Primaria		Matemática	Resuelve problemas de forma, movimiento y localización.	
			Resuelve problemas de forma, movimiento y localización.	
		Comunicación	Lee diversos tipos de textos escritos en su lengua materna.	
Primaria	6.º	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.	
			Resuelve problemas de forma, movimiento y localización.	

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 et al (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 2024, demanda muchos recursos y tiempo.

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; Muñoz, 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).

Muñoz (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. Esta correspondencia de las matrices es validada por los especialistas de las direcciones de línea del Ministerio de Educación.

En los siguientes acápites, 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 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 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 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.

1.2.4 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.

1.2.5 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.

Tabla 1.2 Acomodaciones a las pruebas de la ENLA 2024

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

Tipo de discapacidad		Tipo de acomodación	Descripción
Visual	Ceguera total	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
		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	Hipoacusia/ Sordera	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.	
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

Tipo de discapacidad	Tipo de acomodación	Descripción
Física	Apoyo	Aplicador personalizado para asistirlo en el marcado de las respuestas (si el estudiante lo requería)
	Presentación	Prueba sin adaptaciones en el formato
	Equipos/materiales	Tarjetas de comunicación alternativa (si el estudiante lo requería)
	Asperger	
	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
	Presentación	Prueba con léxico y sintaxis simplificados
	Equipos/materiales	Tarjetas de comunicación alternativa
	Autismo	
Trastorno del espectro autista	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 2024

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)

En la ENLA 2024, se implementaron las siguientes acomodaciones en las pruebas y procesos de aplicación:

1.2.6 Aplicación definitiva

La aplicación definitiva de la ENLA 2024 se llevó a cabo durante los días 19 y 20 de noviembre tanto en 4.º como en 6.º grados de primaria. En 4.º grado de primaria, en el primer día, 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 obedeció a la necesidad de neutralizar los posibles efectos del cansancio de los estudiantes, pues rindieron dos pruebas en un día. En el caso de 6.º grado de primaria, en el primer día, se aplicó la prueba de Lectura y, el segundo día, la de Matemática.

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 2024 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 utilizó 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) permitieron 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 posibilitaron que las formas se adapten al nivel de habilidad de grupos de estudiantes diferentes. En este sentido, se elaboraron formas con un mayor nivel de dificultad o para zona A, las cuales fueron aplicadas a estudiantes de escuelas urbanas, así como formas con menor nivel de dificultad o para zona B, que fueron aplicadas a estudiantes de escuelas rurales y a 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, como ya se había mencionado, estas formas tuvieron 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. Asimismo, este bloque 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 2024

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 4.º grado de primaria

La prueba de Lectura de 4.º grado de primaria de la ENLA 2024 estuvo conformada por 17 textos de diferente tipo, género, formato y complejidad a partir de los cuales

se elaboraron 100 ítems. Dichos textos e ítems se distribuyeron en 20 bloques. De ellos, 10 bloques son comunes y 10 son no comunes. Del grupo de bloques comunes, dos 2 comunes a las formas del día 1 y 4, a las formas del día 2. Asimismo, 4 bloques comunes se distribuyen en diferentes posiciones de las formas. Por otro lado, de los 10 bloques no comunes que se encuentran distribuidos entre las distintas formas, 4 bloques son fijos (2 por día), 4 bloques no son fijos (2 por día) y 2 bloques aparecen solo en una forma (1 por día). Las diferentes combinaciones de los bloques permitieron ensamblar 12 formas en total: 6 formas para escuelas con estudiantes de mayor habilidad o zona A (3 formas por día) y 6 para escuelas con estudiantes de menor habilidad o zona B (3 formas por día). De acuerdo con esta distribución, cada forma y el conjunto de estas reflejan los contenidos (capacidades,

Tabla 1.4 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	
	B03	
	B02	
	B06	
2	B06	24
	B01	
	B03	
	B02	
	B07	
3	B08	24
	B01	
	B03	
	B02	
	B09	
4	B09	24
	B01	
	B04	
	B02	
	B10	
5	B10	24
	B01	
	B04	
	B02	
	B07	

Forma	Bloque	Ítems por forma
6	B07	24
	B01	
	B04	
	B02	
	B05	

Tabla 1.5 Matriz de bloques de la prueba de Lectura de 4.º grado de primaria. Día 2

Forma	Bloque	Ítems por forma
7	B15	24
	B12	
	B11	
	B14	
	B16	
8	B16	24
	B12	
	B11	
	B14	
	B17	
9	B17	24
	B12	
	B11	
	B14	
	B18	
10	B15	24
	B13	
	B11	
	B14	
	B19	
11	B17	24
	B13	
	B11	
	B14	
	B20	

Forma	Bloque	Ítems por forma
12	B20	24
	B13	
	B11	
	B14	
	B18	

tipo de textos, desempeños lectores, 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.

De las 12 formas ensambladas, cada estudiante evaluado se enfrentó a 2 de ellas (una por día de aplicación). En consecuencia, de los 100 ítems que tiene esta prueba, cada estudiante respondió 48.

Diseño de bloques de las pruebas de Lectura de 6.º grado de primaria

La prueba de Lectura de 6.º grado de primaria elaborada para la ENLA 2024 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 (6 comunes a todas las formas y 8 comunes a las formas correspondientes a cada día de aplicación; 4 por día) y 12 bloques son no comunes y se encuentran distribuidos entre las distintas formas. De ellos, 4 bloques son comunes fijos (2 por día) y 12 bloques son no comunes (6 por día) que se encuentran distribuidos entre las distintas formas. Las diferentes combinaciones de los bloques permitieron ensamblar 6 formas para escuelas con estudiantes de mayor habilidad o zona A y 6 para escuelas con

Tabla 1.6 Matriz de bloques de la prueba de Lectura de 6.º grado de primaria. Día 1

Forma	Bloque	Ítems por forma
1	B05	24
	B01	
	B02	
	B06	
	B06	
2	B01	24
	B02	
	B07	
	B07	

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

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.

De las doce formas ensambladas, cada estudiante evaluado se enfrentó a una de ellas (aplicada solo en el primer día). En consecuencia, de los 96 ítems que tiene esta prueba, cada estudiante respondió 24.

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 2024 estuvo conformada por un total de 112 ítems organizados en 30 bloques: 18 bloques comunes y 12 no comunes. Con estos 30 bloques, se ensamblaron 12 formas. Cada forma contiene 25 ítems. Todas las formas están constituidas por 19 ítems comunes organizados en 6 bloques comunes, 2 en posición fija y 4 que varían de posición de una forma a otra; además de 6 ítems no comunes organizados en 2 bloques que varían de posición de una forma a otra. Las diferentes combinaciones de los bloques permitieron ensamblar 6 formas para escuelas con estudiantes de mayor habilidad o

Tabla 1.7 Matriz de bloques de la prueba de Matemática de 4.º grado de primaria.
Día 1

Forma	Bloque	Ítems por forma
1	B01	25
	B13	
	B27	
	B19	
	B25	
	B28	
	B14	
	B02	
2	B02	25
	B14	
	B28	
	B20	
	B25	
	B27	
	B15	
	B03	
3	B03	25
	B15	
	B27	
	B21	
	B25	
	B28	
	B13	
	B01	
4	B04	25
	B13	
	B28	
	B19	
	B25	
	B27	
	B14	
	B05	

Forma	Bloque	Ítems por forma
5	B05	25
	B14	
	B27	
	B20	
	B25	
	B28	
	B15	
	B06	
6	B06	25
	B15	
	B28	
	B21	
	B25	
	B27	
	B13	
	B04	

Tabla 1.8 Matriz de bloques de la prueba de Matemática de 4.º grado de primaria.
Día 2

Forma	Bloque	Ítems por forma
7	B07	25
	B16	
	B29	
	B22	
	B26	
	B30	
	B17	
	B08	
8	B08	25
	B17	
	B30	
	B23	
	B26	
	B29	
	B18	
	B09	

Forma	Bloque	Ítems por forma
9	B09	25
	B18	
	B29	
	B24	
	B26	
	B30	
	B16	
	B07	
10	B10	25
	B16	
	B30	
	B22	
	B26	
	B29	
	B17	
	B11	
11	B11	25
	B17	
	B29	
	B23	
	B26	
	B30	
	B18	
	B12	
12	B12	25
	B18	
	B30	
	B24	
	B26	
	B29	
	B16	
	B10	

zona A y 6 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.

De las doce formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 112 ítems que tiene esta prueba, cada estudiante respondió 50.

Diseño de bloques de las pruebas de Matemática de 6.º grado de primaria

La prueba de Matemática de 6.º grado de primaria elaborada para la ENLA 2024 estuvo conformada por un total de 104 ítems organizados en 34 bloques: seis bloques comunes y 28 no comunes. Con estos 34 bloques, se ensamblaron diez formas. Cada forma contiene 25 ítems. Las formas 1 y 4 están constituidas por 10 ítems comunes organizados en un bloque común en posición fija y 15 ítems no comunes organizados en 7 bloques, 6 que varían de posición de una forma a otra y 1 bloque en posición fija. Las formas 2, 3, 5, 6, 7, 8, 9 y 10 están constituidas por 11 ítems comunes organizados en 2 bloques comunes, uno en posición fija y uno que varía de posición de una forma a otra; además de 14 ítems no comunes organizados en 6

Tabla 1.9 Matriz de bloques de la prueba de Matemática de 6.º grado de primaria. Día 2

Forma	Bloque	Ítems por forma
1	B01	25
	B11	
	B25	
	B19	
	B23	
	B26	
	B12	
2	B02	25
	B03	
	B13	
	B27	
	B20	
	B23	
	B28	
	B14	
	B04	

Forma	Bloque	Ítems por forma
3	B05	25
	B15	
	B29	
	B21	
	B23	
	B30	
	B16	
	B06	
4	B02	25
	B12	
	B26	
	B22	
	B23	
	B27	
	B13	
	B03	
5	B04	25
	B14	
	B28	
	B19	
	B23	
	B29	
	B15	
	B05	
6	B06	25
	B16	
	B30	
	B20	
	B23	
	B25	
	B11	
	B01	
7	B07	25
	B17	
	B31	
	B19	
	B24	
	B32	

Forma	Bloque	Ítems por forma
7	B14	25
	B08	
	B09	
	B18	
	B33	
8	B20	25
	B24	
	B34	
	B16	
	B10	
9	B08	25
	B14	
	B32	
	B21	
	B24	
10	B33	25
	B18	
	B09	
	B10	
	B16	
	B34	
	B22	
	B24	
	B31	
	B17	
	B07	

bloques que varían de posición de una forma a otra. Las diferentes combinaciones de los bloques permitieron ensamblar 6 formas para escuelas con estudiantes de mayor habilidad o zona A y 4 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.

De las diez formas ensambladas, cada estudiante evaluado se enfrentó a una de ellas (aplicada solo en el segundo día). En consecuencia, de los 104 ítems que tiene esta prueba, cada estudiante respondió 25.

Población a evaluar

Capítulo 2

Capítulo 2

Población a evaluar

2.1 Población objetivo

La ENLA 2024 tuvo como población objetivo a los estudiantes que asistieron a clases en 4.º grado de primaria y 6.º grado de primaria de Instituciones Educativas de Educación Básica Regular (EBR).

Se evaluó a los estudiantes de los grados mencionados porque son aquellos en donde se concluye el cuarto y quinto ciclo de la EBR respectivamente, lo que brinda la oportunidad de planificar intervenciones en el sistema para mejorar los grados y ciclos posteriores, en los cuales 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 de esta exclusión 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
4.º grado de primaria	4,1
6.º grado de primaria	3,8

2.2 Marco censal y marco muestral

En la ENLA 2024 se llevó a cabo una evaluación censal en 4.º grado de primaria y una evaluación muestral en 6.º grado de primaria. En este sentido el listado conformado por todos los elementos que potencialmente pueden ser evaluados será denominado marco muestral cuando nos referimos a la evaluación muestral de 6.º grado y será denominado marco censal cuando nos referimos a la evaluación censal de 4.º grado. Ambos marcos son construidos siguiendo la misma metodología, la única diferencia será que en el caso del marco muestral se seleccionará una muestra

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.

probabilística a partir de él mientras que para caso del 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 2023 y la información del Sistema de Información de apoyo a la Gestión de la Institución Educativa (Siagie) del 2023/2024. Anualmente, la Unidad de Estadística Educativa (UEE) del Minedu elabora y difunde el Padrón de Instituciones Educativas con fines estadísticos, el cual es actualizado de manera constante. La información sobre las IE tiene como fuente las operaciones estadísticas que aplican a nivel nacional y los procedimientos asociados a estas operaciones. 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 2024, se aplicó el siguiente procedimiento:

1. Se utilizó el Padrón de Instituciones Educativas para determinar el listado de Instituciones Educativas activas². Se seleccionaron las Instituciones Educativas de primaria de Educación Básica Regular (EBR). A continuación, se excluyeron a aquellas IE que pertenecieran al tipo de atención fortalecimiento en Educación Intercultural Bilingüe (EIB) según el Registro Nacional de Instituciones Educativas del año 2022 que brindan el Servicio de Educación Intercultural Bilingüe (RNIIEE EIB).
2. Se calculó la cantidad de secciones y estudiantes programados. Esta información corresponde al valor registrado en Siagie en 2023. Si una IE no cuenta con información de Siagie se tomó el valor del Censo Educativo 2023.
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 ya que esta es del tipo censal. En este sentido, esta sección solo es aplicable para 6.º grado de primaria.

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 bietápico, donde la unidad primaria de muestreo corresponde a las instituciones educativas 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 conglomerados con probabilidades iguales puede resultar en estimaciones para la media aritmética con una varianza mayor que para un diseño proporcional al tamaño, además de que 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 podría derivar en la selección de muchas escuelas pequeñas que podrían complicar la operación de campo. Finalmente, en la segunda etapa, se realizó una selección aleatoria simple de hasta 2 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 es necesario 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 2024 se realizó una estratificación principalmente porque, en primer lugar, produce un límite más pequeño para el error de estimación que el que se generaría por una muestra aleatoria simple del mismo tamaño (Sheaffer, 2006). En segundo lugar, según un muestreo estratificado nos protegerá de una mala muestra en la que podríamos tener un desbalance en las cantidades seleccionadas 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 característica), por lo que es importante reportar resultados para cada una de ellas. La combinación de estas tres subpoblaciones permite generar los siguientes estratos mutuamente excluyentes y colectivamente exhaustivos donde cada elemento pertenece únicamente a un estrato.

En el caso de 6.º grado de 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).

Para la evaluación muestral de 6.º grado de primaria, además de los estratos antes mencionados se cuenta con 26 dominios que vendrían a ser cada Dirección Regional de Educación (DRE), bajo la consideración de que las IE son administradas por estas instancias y en estas se forman los estratos anteriormente descritos. Por ello, las DRE son combinadas con los estratos antes mencionados, teniendo casos como el de Lima Metropolitana y Callao, que solo consideran 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 es de 123.

Tamaño de muestra

El diseño de muestra utilizado en la UMC es un diseño complejo y como tal, a falta de una fórmula explícita, se recurre al efecto diseño como sugiere Kish (1972). Para el cálculo del tamaño de muestra se utiliza como insumo el cálculo del efecto diseño (DEFF) de evaluaciones previas y se establece 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 calcula el tamaño de muestra para cada DRE y dentro de cada una se estratifica 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 tamaño del mismo.

La afijación de Neyman o 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 2024

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 2024, la ENLA tenía programado evaluar 22 589 IE y aproximadamente 667 402 estudiantes entre todos los grados programados.

El porcentaje de estudiantes evaluados respecto de los programados a nivel nacional en la evaluaciones censales y muestrales varía entre 88% y 98% dependiendo del grado. 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 2024.

Tabla 2.2 Cobertura de la ENLA 4P 2024 a nivel de IE y estudiantes.

4.° grado de primaria	Programados		Evaluados		Cobertura (%)	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Total	18138	557129	18112	532203	99,9	95,5
Amazonas	355	5841	355	5661	100,0	96,9
Áncash	597	17821	597	17247	100,0	96,8
Apurímac	180	5686	180	5595	100,0	98,4
Arequipa	920	26697	920	25971	100,0	97,3
Ayacucho	350	9598	348	9183	99,4	95,7
Cajamarca	1538	24984	1538	24339	100,0	97,4
Callao	444	18717	444	17668	100,0	94,4
Cusco	606	20314	606	19973	100,0	98,3
Huancavelica	217	3955	217	3878	100,0	98,1
Huánuco	557	13157	557	12664	100,0	96,3
Ica	479	18764	479	18000	100,0	95,9
Junín	909	23176	909	22365	100,0	96,5
La Libertad	1345	37202	1345	35641	100,0	95,8
Lambayeque	731	24993	730	23665	99,9	94,7
Lima Metropolitana	3758	153770	3751	145398	99,8	94,6
Lima Provincias	616	19538	616	18770	100,0	96,1
Loreto	841	21461	825	19084	98,1	88,9
Madre de Dios	109	4741	109	4468	100,0	94,2

4.° grado de primaria	Programados		Evaluados		Cobertura (%)	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Moquegua	79	3070	79	3019	100,0	98,3
Pasco	226	4765	226	4537	100,0	95,2
Piura	1372	39852	1372	38401	100,0	96,4
Puno	574	16777	574	16516	100,0	98,4
San Martín	728	18269	728	17614	100,0	96,4
Tacna	149	5854	149	5758	100,0	98,4
Tumbes	135	5279	135	4944	100,0	93,7
Ucayali	323	12848	323	11844	100,0	92,2

Tabla 2.3 Cobertura de la ENLA 6P 2024 a nivel de IE y estudiantes.

6.° grado de primaria	Programado		Evaluado		Cobertura (%)	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Total	4451	110273	4442	105534	99,8	95,7
Amazonas	210	3104	209	2994	99,5	96,5
Áncash	205	5328	205	5138	100,0	96,4
Apurímac	125	3177	125	3109	100,0	97,9
Arequipa	142	4172	142	4081	100,0	97,8
Ayacucho	169	3992	168	3855	99,4	96,6
Cajamarca	376	5586	376	5437	100,0	97,3
Callao	116	4630	116	4420	100,0	95,5
Cusco	153	4263	153	4200	100,0	98,5
Huancavelica	178	2685	178	2634	100,0	98,1
Huánuco	220	4529	220	4311	100,0	95,2
Ica	175	5975	175	5781	100,0	96,8
Junín	227	4872	227	4700	100,0	96,5
La Libertad	182	4289	182	4104	100,0	95,7
Lambayeque	160	5067	160	4823	100,0	95,2
Lima Metropolitana	94	3437	94	3294	100,0	95,8
Lima Provincias	144	3807	144	3665	100,0	96,3
Loreto	246	5869	241	5174	98,0	88,2
Madre de Dios	79	2550	79	2414	100,0	94,7

6.° grado de primaria	Programado		Evaluado		Cobertura (%)	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Moquegua	57	1819	57	1777	100,0	97,7
Pasco	132	2381	132	2244	100,0	94,2
Piura	243	6177	243	5970	100,0	96,6
Puno	204	4563	203	4495	99,5	98,5
San Martín	246	5763	245	5566	99,6	96,6
Tacna	75	2605	75	2566	100,0	98,5
Tumbes	96	3101	96	2904	100,0	93,6
Ucayali	197	6532	197	5878	100,0	90,0

Respecto a los niveles de inferencia de la ENLA 2024 en 4° grado de primaria, dado que es una evaluación censal, se pueden reportar resultados representativos para cualquier nivel de desagregación siempre y cuando se tenga en cuenta la cobertura alcanzada a nivel de estudiantes y escuelas.

En 6° 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.

2. 5 Pesos muestrales ENLA 2024

Los pesos muestrales, aplicados en la ENLA de 6.° grado de primaria, 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 3 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, el cual está dado por la siguiente fórmula:

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

donde:

$p_{f_{kij}}$ = 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.

pseci = peso por sección en la i-ésima IE.

aki = ajuste por estudiantes no evaluados en la k-ésima sección de la i-ésima IE.

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

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 realizaron los siguientes pasos:

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

$$a1_i = \frac{t_i}{s_i}$$

donde:

a1i = Ajuste por estudiantes no evaluados en la i-ésima IE.

ti = total de estudiantes matriculados en la i-ésima IE.

Si = total de estudiantes evaluados en la i-ésima IE.

Ajuste por estudiantes no evaluados en el estrato. Se construyó una estratificación combinando el distrito, la gestión y el área de la escuela.

$$a2_j = \frac{PROG_j}{EVA_AJ_j}$$

Donde:

a2j = Ajuste por estudiantes no evaluados en el estrato j

EVA_AJj = Cantidad de estudiantes evaluados ajustada en el estrato j

PROGj = Cantidad de estudiantes matriculados en el estrato j

Ajuste por no respuesta. Finalmente se obtiene un peso para lectura y otro para matemática, el cual está dado por la siguiente fórmula:

$$AJNR_{ij} = a1_i \times a2_j$$

donde:

AJNRij = Ajuste por no respuesta de la i-ésima IE en el j-ésimo estrato.

a1i = Ajuste por estudiantes no evaluados en la i-ésima IE.

a2j = Ajuste por estudiantes no evaluados en el estrato j

2. 6 Muestra de 2.º de secundaria

En el 2024 por solicitud del Ministerio de Economía y finanzas se programaron dentro del operativo de la ENLA 2024 un total de 817 instituciones educativas y 91 833 estudiantes. Se aplicaron un total de 3 instrumentos relacionados a temas de bienestar y convivencia escolar los cuales fueron un cuestionario al director, un cuestionario al docente tutor y un cuestionario al estudiante. En este grado no se aplicaron pruebas de rendimiento.

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

DRE	Programados		Evaluados		Cobertura (%)	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Amazonas	8	883	8	819	100,0	92,8
Áncash	35	4072	35	3648	100,0	89,6
Apurímac	4	446	4	426	100,0	95,5
Arequipa	24	2335	24	2175	100,0	93,1
Ayacucho	16	1591	14	1229	87,5	77,2
Cajamarca	19	2018	19	1790	100,0	88,7
Callao	43	5094	43	4526	100,0	88,8
Cusco	28	3011	28	2867	100,0	95,2
Huánuco	20	2135	20	1832	100,0	85,8
Ica	18	1994	18	1766	100,0	88,6
Junín	9	818	9	771	100,0	94,3
La Libertad	66	7789	66	6922	100,0	88,9
Lambayeque	30	3367	30	2932	100,0	87,1
Lima Metropolitana	290	33278	290	29448	100,0	88,5
Lima Provincias	17	1611	17	1439	100,0	89,3
Loreto	43	5007	43	3890	100,0	77,7
Madre de Dios	8	1010	8	905	100,0	89,6
Moquegua	6	654	6	606	100,0	92,7

DRE	Programados		Evaluados		Cobertura (%)	
	IE	estudiantes	IE	estudiantes	IE	estudiantes
Piura	49	5997	49	5266	100,0	87,8
San Martín	20	2011	20	1744	100,0	86,7
Tacna	21	1983	21	1868	100,0	94,2
Tumbes	11	1154	11	1031	100,0	89,3
Ucayali	32	3625	32	2786	100,0	76,9

Operativo de campo

Capítulo 3

Capítulo 3

Operativo de campo

Considerando la programación de evaluaciones prevista para el 2024 y el retorno, después de cinco años, a una evaluación censal, fue necesario revisar y replantear las estrategias aplicadas en años anteriores, con el fin de asegurar el cumplimiento de los objetivos de la ENLA 2024. Estas estrategias quedaron plasmadas en las especificaciones técnicas de los Términos de Referencia del convenio y de los servicios requeridos, así como en los manuales de procedimientos elaborados por la UMC.

Como parte del diseño operativo de la ENLA 2024, los procedimientos del operativo de campo se organizaron, implementaron y ejecutaron en base a tres principios fundamentales:

- 1. Confidencialidad:** Garantizar que únicamente el aplicador y los estudiantes tengan acceso a las pruebas.
- 2. Estandarización:** Asegurar que los instrumentos se apliquen en las mismas condiciones y siguiendo los procedimientos establecidos.
- 3. Probidad:** Fomentar la rectitud, honradez y honestidad en las acciones de todos los actores involucrados en el proceso de evaluación.

Dentro de estas estrategias, se consideró fundamental continuar contando con el INEI como operador logístico, debido a su amplia experiencia técnica en la aplicación a gran escala de instrumentos de evaluación estandarizada. Desde el año 2007, mediante convenios de cooperación interinstitucional, el INEI ejecuta la aplicación de las Evaluaciones Nacionales de Logros de Aprendizaje.

A lo largo de todas las ediciones de estas evaluaciones, el INEI y el Minedu han trabajado de manera conjunta, realizando los esfuerzos necesarios para garantizar la correcta ejecución de la aplicación.

Las responsabilidades asignadas a cada institución durante este año fueron las siguientes:

Responsabilidades del INEI

- 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).
- Resguardar los instrumentos y documentos de aplicación desde su llegada a cada local de la sede de jurisdicción hasta su recojo.
- 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:

- Imprimir, inventariar y modular los instrumentos de aplicación (cuadernillos de pruebas, cuestionarios, fichas ópticas y documentos auxiliares).
- Imprimir y modular los materiales de capacitación.
- Distribuir los materiales de capacitación.
- Distribuir y replegar los materiales de aplicación.
- Conseguir los locales en las sedes jurisdiccionales e implementarlos con mobiliarios y servicios.
- Implementar equipos informáticos en las sedes jurisdiccionales a nivel nacional.

Asimismo, el Minedu acompañó y monitoreó todas las etapas de la implementación de la evaluación, con el propósito de verificar el cumplimiento de los procesos y procedimientos establecidos. Para ello, desplegó personal en las sedes jurisdiccionales, encargado de realizar labores de coordinación y seguimiento, acompañar a la Red Administrativa (RA) en todas las etapas de ejecución de la evaluación y brindar apoyo en la atención de incidentes que surgieran en las sedes. De igual manera, supervisó las actividades desarrolladas por los contratistas responsables de los procesos de impresión, distribución, acondicionamiento de locales e instalación de equipos informáticos, con el objetivo de garantizar el cumplimiento de los Términos de Referencia y la ejecución oportuna de las actividades programadas.

3.1 Organización territorial

Considerando el principio de accesibilidad, el número de instituciones educativas, los grados y los operativos previstos, así como con el fin de facilitar la ejecución de la evaluación, el país fue organizado territorialmente en 37 sedes jurisdiccionales regionales, dentro de las cuales se establecieron e implementaron 267 sedes jurisdiccionales.

Cabe mencionar que por solicitud del Ministerio de Economía y finanzas participaron del operativo de la ENLA 2024 un total de 815 instituciones educativas y 80 676 estudiantes. Se aplicaron un total de 3 instrumentos relacionados a temas de bienestar y convivencia escolar los cuales fueron un cuestionario al director, un cuestionario al docente tutor y un cuestionario al estudiante. En este grado no se aplicaron pruebas de rendimiento.

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

Sede Regional	4P Censal	6P Muestral regional	2S MEF	Total general
AMAZONAS	748	318	16	1082
ANCASH-CHIMBOTE	328	117	23	468
ANCASH-HUARAZ	276	89	12	377
APURIMAC	182	127	4	313
AREQUIPA-NORTE	471	66	10	547
AREQUIPA-SUR	458	77	14	549
AYACUCHO	337	138	17	492
CAJAMARCA	1203	274	11	1488
CALLAO	452	118	43	613
CUSCO	573	147	27	747
HUANCAVELICA	214	167	--	381
HUANUCO	517	199	20	736
ICA	580	225	18	823
JUNIN	942	227	9	1178
LA LIBERTAD-NOR OESTE	532	69	42	643
LA LIBERTAD-SUR ESTE	814	111	24	949
LAMBAYEQUE	754	160	30	944
LIMA MET 1	713	24	45	782
LIMA MET 2	661	11	51	723
LIMA MET 3	380	7	15	402
LIMA MET 4	672	20	50	742
LIMA MET 5	562	18	61	641
LIMA MET 6	658	27	44	729
LIMA MET 7	343	10	24	377
LIMA-CAÑETE	200	49	8	257

Sede Regional	4P Censal	6P Muestral regional	2S MEF	Total general
LIMA-HUACHO	358	76	9	443
LORETO	618	171	32	821
MADRE DE DIOS	108	78	8	194
MOQUEGUA	79	57	6	142
PASCO	222	131	--	353
PIURA	1402	249	50	1701
PUNO	597	204	--	801
SAN MARTIN-MOYOBAMBA	341	120	7	468
SAN MARTIN-TARAPOTO	569	180	21	770
TACNA	153	75	21	249
TUMBES	132	90	10	232
UCAYALI	454	245	35	734
Total general	18603	4471	817	23891

3.2 Conformación de la Red Administrativa de campo

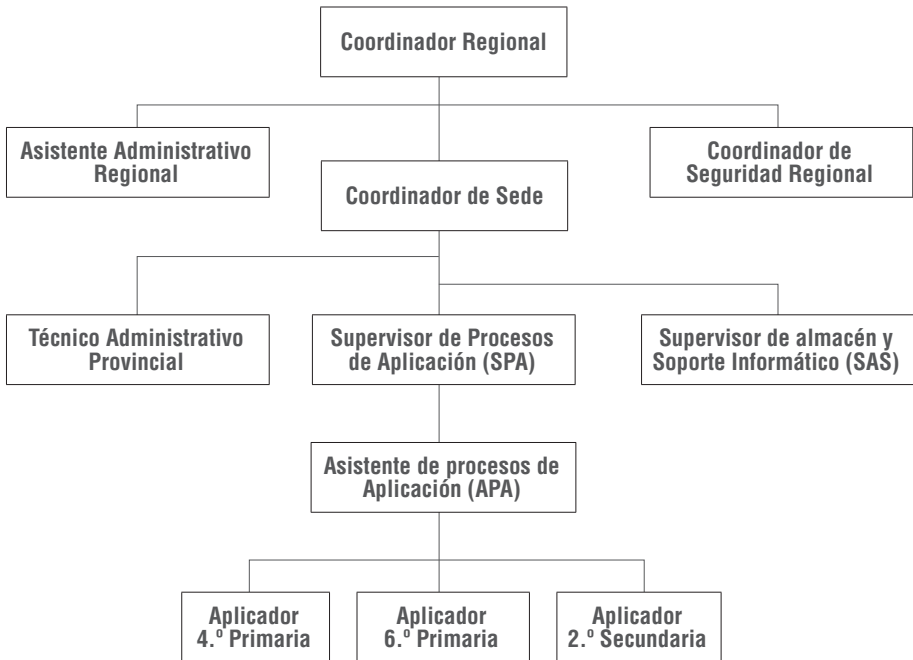
El personal necesario para ejecutar el operativo estuvo organizado en una Red Administrativa (RA) de campo conformada por los cargos mostrados en la tabla 3.2.

Tabla 3.2 Red Administrativa de campo

Nivel	Cargo en la RA de campo	Total RA 2024
Nivel IA	Supervisores nacionales	60
Nivel IB	Coordinadores regionales (CR)	37
Nivel IB-1	Asistente administrativo regional	37
Nivel IB-2	Coordinador de seguridad regional	37
Nivel IC	Coordinadores de sede	267
Nivel IC-1	Técnicos administrativos (incluye todas las sedes jurisdiccionales)	267
Nivel IIB	Supervisores de almacén y soporte informático (SAS)	267
Nivel IIA	Supervisor de procesos de aplicación (SPA)	376
Nivel III	APA	2406
	Aplicadores 4P	38519
Nivel IV	Aplicadores 6P	7732
	Aplicadores 2S	817
	Aplicadores de apoyo a estudiantes con discapacidad (aplicador ED)	2460

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 de campo ENLA 2024



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

- Los supervisores nacionales aseguran el cumplimiento de los procesos y procedimientos de la Evaluación, supervisan que se garanticen las condiciones materiales necesarias para la evaluación en cada sede jurisdiccional y monitorea la consecución de personal en los diferentes niveles de la Red Administrativa de la región asignada.
- Los coordinadores regionales 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 (sede regional). Asimismo, son responsables de garantizar las condiciones materiales necesarias para el operativo en su región.
- Los coordinadores de sede organizan, gestionan y dirigen la Evaluación en su sede jurisdiccional. Garantizan las condiciones materiales necesarias para la Evaluación. Son responsables de la convocatoria, selección y capacitación de los Aplicadores. Supervisan el cumplimiento de los procedimientos y monitorean la labor de todos los miembros de la RA de su provincia.

- Los coordinadores de seguridad regional tienen la responsabilidad de supervisar la eficacia de las medidas preventivas frente a riesgos que puedan afectar la integridad de la Red Administrativa (RA) y de los locales bajo su jurisdicción. Asimismo, velan por el resguardo de la confidencialidad de los instrumentos de evaluación durante el desarrollo del operativo.
- Los supervisores de procesos de aplicación se encargan de apoyar en la supervisión e implementación del proceso de convocatoria, selección y capacitación de Asistentes de procesos de aplicación. Verifican que los APA y aplicadores cumplan con lo señalado en los manuales y protocolos.
- Los supervisores de almacén y soporte informático son responsables de gestionar los sistemas de inventario y la actualización de la información de las instituciones educativas. Además, se encargan de la recepción, almacenamiento, inventario y despacho del material de aplicación, así como del uso adecuado de los sistemas de información implementados para estas tareas.
- Los asistentes de procesos de aplicación son responsables de realizar el contacto previo con las IE asignadas, llevar a cabo la selección y capacitación de los aplicadores, velar por el cumplimiento de los procedimientos y supervisar la aplicación.
- Los aplicadores son responsables de la adecuada aplicación de los instrumentos en las IE siguiendo los procedimientos descritos en los manuales.

Para cada uno de los niveles de la RA, el Minedu entregó manuales de aplicación y de funciones. En total, se elaboraron 16 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 de los sistemas
- Manual del asistente de procesos de aplicación
- Protocolo del asistente de procesos de aplicación secundaria
- Manual del aplicador 4.º grado de primaria
- Manual del aplicador 6.º grado de primaria
- 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 determinó, en los Términos de Referencia del convenio suscrito con el INEI, los requisitos profesionales necesarios para cada uno de los puestos de la RA de campo. Asimismo, tomando como referencia la experiencia obtenida en operativos anteriores y el análisis de los desafíos identificados, se establecieron los siguientes perfiles para el personal aplicador:

Perfil del aplicador

I. Personal con experiencia en aplicación de evaluaciones estandarizadas:

OPCIÓN 1

- Personal con, mínimo, una (01) experiencia demostrada, en la aplicación de evaluaciones estandarizadas y que su desempeño no haya sido observado.
- Ser mínimo estudiante universitario o de instituto pedagógico o de carrera técnica.

II. Personal sin experiencia en aplicación de evaluaciones estandarizadas:

OPCIÓN 2

Perfil académico:

- Egresado universitario o egresado técnico o egresado pedagógico, excepto carreras vinculadas al cuidado personal u oficios.

Experiencia sustentada:

- Como facilitador, capacitador o educador (sin carga docente en el sector estatal); o
- en la aplicación de instrumentos de recojo de información; o
- en participación en actividades con niñas, niños y adolescentes o en proyectos sociales.

OPCIÓN 3

Perfil académico:

- Mínimo estudiante universitario de los últimos dos años; o estudiantes de institutos pedagógicos o técnicos del último año.

Experiencia sustentada:

- Como facilitador, capacitador o educador (sin carga docente en el sector estatal); o
- en la aplicación de instrumentos de recojo de información; o
- en participación en actividades con niñas, niños y adolescentes o en proyectos sociales.

Adicionalmente para la opción 2 y 3:

- 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 y para viajar o trasladarse y/o pernoctar al lugar que se le designe.
- Obligatorio: que disponga de teléfono móvil.

Perfil adicional de aplicador (previa coordinaciones entre el INEI y la UMC):

I. Personal con experiencia en aplicación de evaluaciones estandarizadas:

OPCIÓN 1: no necesita cumplir con otros requisitos del perfil

- Con experiencia como Aplicador en al menos uno (1) de los operativos de la UMC de los años 2015, 2016, 2018, 2019, 2022 o 2023 y que su desempeño no haya sido observado.

II. Personal sin experiencia en aplicación de evaluaciones estandarizadas:

OPCIÓN 2

Perfil académico:

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

Experiencia laboral:

Con práctica sustentada:

- Participación en actividades con niñas, niños y adolescentes en aula, iglesia o en proyectos sociales; o
- participación demostrada en proyectos de salud, educación, o alguna otra área social; o
- facilitación de actividades educativas o sociales.

OPCIÓN 3

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

- Participación en actividades con niñas, niños y adolescentes en aula, iglesia 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 por considerar para las OPCIONES 2 y 3:

- Edad entre 18 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 y para viajar al lugar que se le designe.
- Imprescindible: 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):

- Estudiante universitario de carreras de educación especial de los dos últimos años; o
- egresado universitario/o titulado técnico con especialidad o diplomado en educación inclusiva; 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, fisioterapia y rehabilitación, terapia de lenguaje, terapia ocupacional o carreras afines.

Además, deberá contar con:

- Una (1) experiencia de trabajo con niños y niñas, o adolescentes o con 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; deseable manejo del sistema de comunicación por intercambio de tarjetas.

Para la ENLA 2024, se planificaron jornadas de capacitación tanto centralizadas — en Lima y en las capitales regionales— como descentralizadas, llevadas a cabo en cada sede jurisdiccional. Únicamente la capacitación dirigida a los aplicadores se realizó en sus respectivas sedes jurisdiccionales.

El propósito de estas capacitaciones fue fortalecer las competencias de los equipos responsables, familiarizándolos con los procedimientos e instrumentos de evaluación, y garantizar que todos recibieran las mismas directrices a nivel nacional. Las capacitaciones dirigidas a los niveles superiores de la Red Administrativa (RA) se enfocaron en afianzar los procesos clave requeridos para el desarrollo integral del operativo, de acuerdo con las funciones específicas asignadas a cada nivel. En el caso de los aplicadores, se capacitó a los postulantes en la administración estandarizada de los instrumentos, empleando para ello el Manual del Aplicador y los siguientes recursos diseñados específicamente para este fin:

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

La capacitación dirigida a los aplicadores tuvo una duración de dos días y se desarrolló en grupos de hasta 30 participantes por aula. Aquellos aplicadores asignados a atender estudiantes con discapacidad participaron primero en la capacitación general junto con los demás aplicadores, y posteriormente recibieron una jornada adicional de medio día, específica según el tipo de discapacidad que debían 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 para asegurar la estandarización de los procedimientos.

3.4 Características de los instrumentos y documentos utilizados

Se utilizaron los siguientes instrumentos y documentos de aplicación:

4.º grado de primaria

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

6.º grado de primaria

- Cuadernillos integrados de pruebas (un cuadernillo por día de evaluación)
- Ficha óptica de asistencia y respuestas (FOAR)
- Cuestionario al estudiante día 1
- Cuestionario al estudiante día 2
- Cuestionario a la familia

- Carta al director
- Cuestionario al director
- Carta al docente
- Compromiso de confidencialidad del docente
- Cuestionario al docente de comunicación
- Cuestionario al docente de matemática
- Cuestionario al docente tutor
- Ficha del aplicador líder
- Registro de préstamo
- Ficha de verificación

2.º grado de secundaria

- Cuestionario al estudiante
- Registro de estudiantes
- Carta al director
- Cuestionario al Director
- Carta al docente
- Cuestionario al docente tutor
- Compromiso de confidencialidad para el docente
- Ficha del aplicador de sección
- Ficha del aplicador líder
- Registro de préstamo
- Ficha de verificación

Las cajas con los instrumentos de evaluación fueron trasladadas a las sedes jurisdiccionales con un mínimo de seis días de anticipación a las fechas de aplicación. Esta medida permitió que los aplicadores asignados a zonas de difícil acceso pudieran movilizarse con el tiempo necesario y garantizar su presencia en las instituciones educativas en la fecha programada. Una vez recibidas en cada sede, las cajas fueron registradas mediante inventario y almacenadas en el almacén

correspondiente. Estos espacios permanecieron bajo llave y contaron con resguardo privado, siendo custodiados por la Red Administrativa de campo (RA) hasta el repliegue del material hacia Lima Metropolitana para su procesamiento.

3.5 Proceso de aplicación

La duración y las fechas de aplicación de cada operativo de la ENLA 2024 se muestran en la tabla 3.3:

Tabla 3.3 Duración y fechas de aplicación de la ENLA

Grado	Denominación del instrumento	Duración de la aplicación	Fechas de aplicación	Criterio para asignar aplicadores
4.ºP	Lectura	2 días	Del 19 de noviembre al 20 de noviembre	1 aplicador líder para cada IE con pernocte con una sola sección programada en toda la IE
				Un aplicador por sección programada
	Matemática			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
6.ºP	Lectura	2 días	Del 19 de noviembre al 20 de noviembre	1 aplicador líder para cada IE con pernocte con una sola sección programada en toda la IE
				Un aplicador por sección programada
	Matemática			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
2.ºS	Cuestionario Factores Asociados	1 día	Del 19 de noviembre al 21 de noviembre	1 aplicador aplica a una o más secciones dentro de una misma IE.

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

Aplicación primaria (4.° y 6.° 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 cada día de aplicación al local de 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 responsable de la sección 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 y siguió los procedimientos respectivos.
- 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

4.° grado de primaria		6.° grado de primaria	
Primer día	Segundo día	Primer día	Segundo día
Lectura (60 m)	Matemática (60 m)	Lectura (60 m)	Matemática (60 m)
Descanso (30 m)	Descanso (30 m)	Descanso (30 m)	Descanso (30 m)
Matemática (60 m)	Lectura (60 m)	Cuestionario al estudiante día 1 (30 m)	Cuestionario al estudiante día 2 (25 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.

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.

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)

Para este grado, se asignó un único aplicador por institución educativa (IE), quien fue responsable de administrar los cuestionarios a todas las secciones programadas en un plazo máximo de tres (3) días. Esta planificación se estableció en función de las coordinaciones previas realizadas entre el asistente de procesos de aplicación y el director de la IE.

I. Recepción de materiales.

- a. En IE sin pernoctación: Los aplicadores acudieron diariamente a la sede para recoger los cuestionarios.
- b. En IE con pernoctación: Los aplicadores se desplazaron con anticipación 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 cuestionarios, 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 cuestionarios siguiendo la secuencia y los tiempos indicados en la siguiente tabla:

Tabla 3.5 Orden y duración de la sesión de aplicación de 2.º grado de secundaria

Grado y operativo	Duración de la sesión
2.º grado de secundaria	Cuestionario (30 m)

- V. Después de la aplicación en cada sección, en un ambiente privado, el aplicador revisó y organizó todos los instrumentos que debían ser devueltos al Minedu.
- VI. Al finalizar cada día de aplicación, en la dirección, 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.
- VII. Al concluir las actividades en la institución educativa, el aplicador se trasladó a la sede, donde el asistente de procesos de aplicación verificó el material entregado, considerando: la cantidad de documentos devueltos y la calidad de la información recogida.
- VIII. Posteriormente, los asistentes de procesos, junto con los supervisores de almacén y soporte informático, realizaron un inventario detallado de los paquetes con instrumentos. Estos fueron embalados nuevamente en sus cajas originales para su retorno.

Procesamiento de datos y análisis psicométrico

Capítulo 4

Capítulo 4

Procesamiento de datos y análisis psicométrico

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 cuadernillos de pruebas, 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). Las fichas ópticas son almacenadas de forma ordenada, de tal manera que se pueda acceder a ellas en caso sea requerido.
- 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.

Es importante señalar que los instrumentos aplicados en 2.º grado de secundaria fueron entregados al MEF luego de la depuración y armada de base de datos. No se realizó en la UMC ningún análisis de esa información.

4.2 Estrategias de análisis psicométrico

4.2.1 Modelo Rasch para ítems dicotómicos

Para la evaluación nacional de logros de aprendizaje (ENLA) 2024 se utilizó el modelo Rasch para datos 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 como 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

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.

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 la curva característica del ítem (CCI), la cual brinda información concreta sobre la probabilidad de respuesta de una persona ante dicho ítem. Al trazar esta curva, en caso de tener un ítem dicotómico, se establece las siguientes relaciones:

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

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

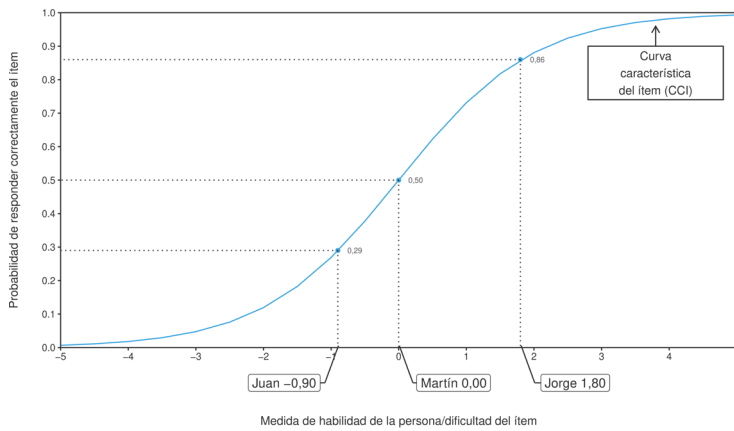
$$\theta = \beta; p(X = 1|\theta, \beta) = 0,5$$

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:

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

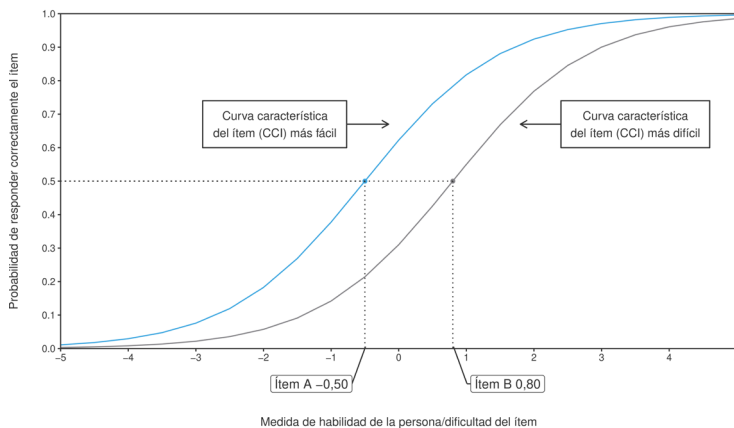
Donde θ representa el rasgo latente de cada persona n , y β representa la medida de dificultad de cada ítem i . Esta relación entre la habilidad de una persona y la dificultad de un ítem se muestra en la figura 4.1.

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 esta 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:

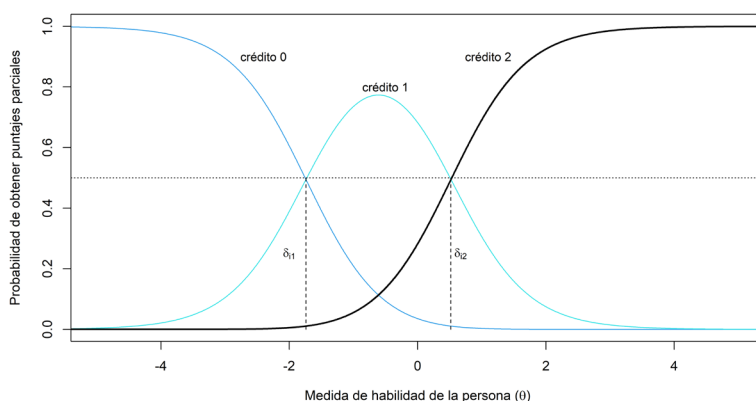
$$\sqrt{\frac{196}{4}} = x$$

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

$$\begin{aligned}\delta_{ij} &= D_i + F_{i1} \\ &\text{y} \\ \delta_{ij} &= D_i + F_{i2}\end{aligned}$$

En este ejemplo, el primer umbral con un valor de $-1,74$ representa la medida en la cual es igualmente probable recibir 0 o 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:

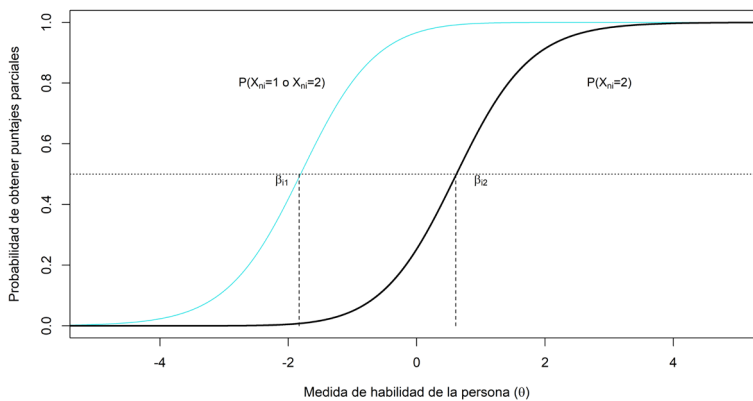
$$P(X_{ni} = 0) = \frac{\exp(0)}{\exp(0) + \exp(0 + (\theta_n - \delta_{i1})) + \exp(0 + (\theta_n - \delta_{i1}) + (\theta_n - \delta_{i2}))}$$

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

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

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 en ítems alejados 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 $\sum_{n=1}^N R_{ni}^2 = VAR(P_{ni})$, 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 basa 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 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 y Matemática

4.3.1 Calibración de los ítems

En las tablas a continuación, 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 Evaluación Nacional de Logros de Aprendizaje (ENLA) 2024. En el caso de los ítems de créditos parciales se presenta los resultados para cada umbral. Las tablas 4.1 y 4.2 presentan las medidas para los ítems de las pruebas de cuarto grado de primaria y las tablas 4.3 y 4.4 contienen información para los ítems de las pruebas de sexto grado de primaria.

Tabla 4.1 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	B15	1	0,896	0,006	1,03	1,11	0,40	0,35	4
CT002	B15	3	0,824	0,006	0,99	1,07	0,43	0,36	4
CT003	B15	4	-0,290	0,005	0,90	0,86	0,51	0,60	3
CT004	B15	5	-1,462	0,007	0,98	0,97	0,37	0,81	2
CT005	B07	1	-1,375	0,006	0,94	0,90	0,43	0,79	2
CT006	B07	2	-0,053	0,005	1,07	1,09	0,40	0,54	3
CT007	B07	3	-0,971	0,005	0,86	0,73	0,53	0,72	2
CT008	B07	4	-0,426	0,005	1,06	1,11	0,39	0,62	3
CT009	B05	5	-0,786	0,006	0,91	0,85	0,49	0,70	3
CT010	B07	5	-1,137	0,006	1,04	1,16	0,36	0,75	2
CT011	B17	2	-0,312	0,004	0,99	0,98	0,44	0,63	3
CT012	B17	3	-0,115	0,004	0,98	0,96	0,46	0,59	3
CT013	B17	5	-1,051	0,005	0,97	0,99	0,41	0,76	2
CT014	B17	4	-0,494	0,004	1,05	1,04	0,39	0,66	3
CT015	B07	6	-0,782	0,005	0,93	0,86	0,49	0,69	3
CT016	B17	1	-1,444	0,005	0,97	0,9	0,39	0,82	2
CT017	B17	6	0,204	0,004	1,04	1,06	0,41	0,52	4
CT018	B13	2	-0,899	0,007	1,06	1,06	0,39	0,62	3
CT019	B20	1	-2,571	0,012	0,90	0,75	0,38	0,88	1
CT020	B20	2	-2,566	0,012	0,86	0,62	0,42	0,88	1
CT021	B20	3	-1,961	0,010	0,87	0,75	0,46	0,81	2
CT022	B16	1	0,420	0,004	1,04	1,07	0,39	0,48	4
CT023	B16	2	0,332	0,004	0,96	0,95	0,47	0,50	4
CT024	B16	3	0,080	0,004	0,87	0,83	0,54	0,56	3
CT025	B16	5	1,190	0,005	1,13	1,31	0,28	0,32	4
CT026	B16	6	0,037	0,004	1,04	1,07	0,38	0,57	3
CT027	B14	5	-0,760	0,003	0,88	0,79	0,52	0,70	3
CT028	B01	5	-1,119	0,004	0,98	1,04	0,40	0,76	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT029	B16	4	-1,033	0,005	0,84	0,71	0,51	0,78	2
CT030	B05	1	0,144	0,005	0,96	0,95	0,48	0,51	4
CT031	B05	2	-1,043	0,006	0,99	1,03	0,39	0,75	2
CT032	B05	6	-1,330	0,006	0,87	0,73	0,48	0,80	2
CT033	B05	3	-0,949	0,006	0,92	0,85	0,47	0,73	2
CT034	B05	4	-0,548	0,006	0,95	0,93	0,46	0,66	3
CT035	B18	1	-1,452	0,007	0,81	0,62	0,53	0,81	2
CT036	B18	2	-0,929	0,006	0,89	0,81	0,50	0,73	2
CT037	B18	3	-0,304	0,006	0,94	0,92	0,49	0,61	3
CT038	B18	4	-0,739	0,006	0,83	0,74	0,56	0,70	3
CT039	B18	5	0,468	0,005	1,02	1,06	0,44	0,45	4
CT040	B18	6	0,266	0,005	1,24	1,33	0,27	0,49	4
CT041	B09	1	-1,420	0,007	0,89	0,82	0,46	0,81	2
CT042	B09	5	-0,957	0,006	0,87	0,76	0,52	0,74	2
CT043	B09	2	-2,021	0,008	0,88	0,70	0,42	0,88	2
CT044	B09	4	-2,179	0,008	0,93	0,92	0,35	0,90	1
CT045	B10	2	-0,451	0,008	0,90	0,86	0,52	0,52	3
CT046	B10	3	-0,612	0,008	0,92	0,86	0,51	0,56	3
CT047	B10	4	0,077	0,009	0,99	1,00	0,46	0,41	3
CT048	B10	5	0,353	0,009	1,09	1,13	0,38	0,35	4
CT049	B04	2	-0,673	0,007	0,91	0,88	0,51	0,57	3
CT050	B02	1	0,128	0,003	0,98	0,97	0,46	0,52	3
CT051	B02	2	0,129	0,003	0,97	0,96	0,47	0,52	3
CT052	B03	2	0,981	0,004	1,09	1,19	0,34	0,36	4
CT053	B02	3	0,577	0,003	1,06	1,12	0,39	0,42	4
CT054	B02	4	1,896	0,084	0,90	0,76	0,45	0,19	4
CT055	B19	1	-0,739	0,012	1,01	1,01	0,42	0,59	3
CT056	B19	2	-0,658	0,012	0,86	0,81	0,54	0,57	3
CT057	B19	3	0,215	0,012	1,08	1,13	0,38	0,38	4
CT058	B19	4	-0,074	0,012	1,06	1,11	0,39	0,44	3
CT059	B19	5	-0,898	0,012	0,95	0,90	0,46	0,62	3
CT060	B19	6	-0,474	0,012	0,97	0,99	0,46	0,53	3
CT061	B09	3	0,371	0,005	1,03	1,06	0,42	0,47	4
CT062	B09	6	-1,754	0,007	0,87	0,74	0,45	0,85	2
CT063	B08	1	0,219	0,006	0,96	0,95	0,47	0,53	4
CT064	B08	2	1,028	0,006	1,09	1,22	0,34	0,36	4
CT065	B08	3	0,412	0,006	1,11	1,17	0,34	0,49	4

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT066	B08	4	-0,438	0,006	1,03	1,02	0,40	0,67	3
CT067	B08	5	0,310	0,006	1,02	1,03	0,42	0,51	4
CT068	B08	6	0,364	0,006	1,03	1,06	0,41	0,50	4
CT069	B06	2	-0,458	0,004	0,98	0,93	0,42	0,67	3
CT070	B06	3	1,266	0,005	1,10	1,30	0,30	0,31	4
CT071	B06	4	-0,393	0,004	0,99	1,02	0,41	0,66	3
CT072	B11	4	1,563	0,078	0,96	0,84	0,44	0,24	4
CT073	B15	6	0,494	0,005	1,05	1,10	0,40	0,43	4
CT074	B06	1	-0,290	0,004	0,97	0,95	0,44	0,64	3
CT075	B06	5	0,275	0,004	1,05	1,06	0,38	0,52	4
CT076	B06	6	0,153	0,004	1,09	1,11	0,35	0,54	4
CT077	B10	1	0,757	0,009	1,09	1,21	0,35	0,28	4
CT078	B10	6	-0,112	0,009	0,95	0,94	0,49	0,45	3
CT079	B20	5	0,077	0,009	1,11	1,20	0,37	0,41	3
CT080	B20	4	0,000	0,009	1,03	1,07	0,44	0,43	3
CT081	B20	6	-	-	-	-	-	-	-
CT082	B04	1	-0,787	0,007	0,93	0,89	0,49	0,60	3
CT083	B03	1	0,816	0,004	1,13	1,23	0,31	0,40	4
CT084	B12	2	1,062	0,004	1,06	1,14	0,36	0,35	4
CT085	B13	1	-1,850	0,008	0,91	0,86	0,43	0,79	2
CT086	B15	2	0,905	0,006	1,05	1,18	0,38	0,35	4
CT087	B14	4	-1,314	0,004	0,91	0,91	0,44	0,79	2
CT088	B14	3	-0,308	0,003	0,86	0,80	0,55	0,61	3
CT089_1	B14	6	-0,248	0,050	0,86	0,81	0,65	0,84	3
CT089_2	B14	6	1,544	0,050	0,86	0,81	0,65	0,84	4
CT090	B01	1	0,126	0,003	1,08	1,12	0,38	0,52	3
CT091	B01	2	-1,374	0,004	0,87	0,78	0,47	0,8	2
CT092	B01	3	0,527	0,003	1,15	1,22	0,32	0,43	4
CT093	B01	6	0,755	0,070	1,06	1,07	0,39	0,39	4
CT094	B01	4	-1,347	0,004	0,98	1,01	0,38	0,80	2
CT095	B12	1	0,653	0,003	1,00	1,02	0,43	0,43	4
CT096	B11	2	0,468	0,003	1,07	1,11	0,39	0,44	4
CT097	B11	1	-0,291	0,003	1,24	1,44	0,23	0,61	3
CT098	B11	3	-0,017	0,003	1,06	1,08	0,40	0,55	3
CT099	B14	1	-0,774	0,003	0,82	0,74	0,56	0,70	3
CT100	B14	2	-0,304	0,003	0,94	0,92	0,49	0,61	3

Tabla 4.2 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	B21	3	-1,138	0,006	0,90	0,90	0,52	0,71	2
MA002	B19	6	-0,345	0,006	0,83	0,77	0,60	0,58	3
MA003	B17	2	0,380	0,004	0,79	0,73	0,63	0,45	4
MA004	B22	7	-1,466	0,006	1,03	1,04	0,41	0,76	2
MA005	B22	3	-1,896	0,007	0,83	0,67	0,52	0,81	2
MA006	B18	1	-1,160	0,004	0,83	0,70	0,57	0,72	2
MA007	B22	1	-0,784	0,006	0,89	0,81	0,55	0,65	3
MA008	B21	5	-2,078	0,007	0,85	0,66	0,50	0,83	1
MA009	B18	3	-1,726	0,005	0,99	1,19	0,41	0,79	2
MA010	B14	3	-0,875	0,004	0,90	0,83	0,54	0,67	3
MA011	B17	3	-1,206	0,004	0,89	0,84	0,52	0,72	2
MA012	B16	1	-1,162	0,004	1,02	0,97	0,45	0,72	2
MA013	B24	4	0,731	0,006	0,88	0,88	0,56	0,40	4
MA014	B18	2	0,179	0,004	1,15	1,29	0,39	0,49	3
MA015	B14	1	-1,836	0,005	0,97	0,91	0,43	0,81	2
MA016	B23	2	0,312	0,006	0,94	1,00	0,52	0,47	3
MA017	B24	3	0,192	0,006	1,02	1,04	0,48	0,49	3
MA018	B13	2	-1,953	0,005	0,92	0,93	0,44	0,82	2
MA019	B21	2	0,589	0,006	0,80	0,77	0,62	0,42	4
MA020	B20	1	-1,791	0,007	0,94	0,83	0,46	0,80	2
MA021	B24	8	-0,373	0,006	0,99	0,98	0,50	0,59	3
MA022	B21	1	-1,320	0,006	0,82	0,69	0,57	0,74	2
MA023	B15	2	-1,255	0,004	0,85	0,74	0,55	0,73	2
MA024	B20	2	0,133	0,006	0,94	0,92	0,53	0,5	3
MA025	B22	2	-0,065	0,006	0,85	0,83	0,59	0,53	3
MA026	B21	7	-0,246	0,006	0,95	0,94	0,53	0,57	3
MA027	B20	5	-1,753	0,007	0,85	0,81	0,51	0,80	2
MA028	B17	1	-2,298	0,005	0,91	0,79	0,43	0,86	1
MA029	B13	1	-1,769	0,005	0,88	0,78	0,50	0,80	2
MA030	B24	7	-2,047	0,007	0,87	0,83	0,48	0,83	1
MA031	B16	3	-1,256	0,004	0,94	0,86	0,49	0,73	2
MA032	B11	1	-2,249	0,010	1,04	1,24	0,37	0,77	1
MA033	B23	8	-1,099	0,006	1,04	1,10	0,42	0,71	2
MA034	B11	2	-2,250	0,010	0,87	0,76	0,50	0,77	1
MA035	B24	6	-2,210	0,007	0,94	0,91	0,42	0,85	1

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA036	B23	1	-2,173	0,007	0,89	0,80	0,45	0,84	1
MA037	B06	3	-1,098	0,009	1,10	1,14	0,42	0,59	2
MA038	B01	1	0,183	0,004	1,26	1,39	0,28	0,52	3
MA039	B06	2	-1,712	0,009	1,02	1,03	0,44	0,69	2
MA040	B22	8	-1,017	0,006	1,04	1,13	0,42	0,69	3
MA041	B22	6	0,011	0,006	1,19	1,30	0,35	0,52	3
MA042	B23	4	-2,489	0,008	1,01	1,29	0,32	0,88	1
MA043	B20	6	-2,311	0,008	0,97	1,07	0,38	0,86	1
MA044	B19	4	0,359	0,006	1,04	1,07	0,46	0,46	4
MA045	B12	3	-0,985	0,009	1,08	1,13	0,44	0,57	3
MA046	B23	3	1,329	0,006	1,07	1,30	0,41	0,30	4
MA047	B14	2	0,407	0,004	1,00	1,05	0,48	0,45	4
MA048	B20	4	0,698	0,006	1,09	1,22	0,42	0,40	4
MA049	B03	2	1,504	0,005	1,01	1,06	0,44	0,30	4
MA050	B24	2	-0,641	0,006	1,03	1,04	0,46	0,63	3
MA051	B09	3	2,081	0,005	0,94	1,13	0,45	0,21	4
MA052	B12	1	-2,507	0,010	0,94	0,98	0,42	0,80	1
MA053	B24	5	0,803	0,006	1,16	1,31	0,37	0,38	4
MA054	B19	5	-1,358	0,006	0,99	1,04	0,44	0,74	2
MA055	B19	7	0,245	0,006	1,08	1,15	0,43	0,48	3
MA056	B02	3	1,384	0,005	1,06	1,24	0,40	0,31	4
MA057	B07	3	-0,532	0,005	0,97	0,97	0,48	0,65	3
MA058	B21	8	-0,589	0,006	0,84	0,78	0,59	0,62	3
MA059	B26	3	-0,239	0,003	1,00	1,03	0,49	0,56	3
MA060	B19	3	0,102	0,006	1,00	1,04	0,48	0,50	3
MA061	B20	8	-1,524	0,006	0,92	0,90	0,48	0,77	2
MA062	B01	2	0,736	0,004	1,22	1,41	0,30	0,42	4
MA063	B08	3	-	-	-	-	-	-	-
MA064	B02	1	0,067	0,004	1,05	1,07	0,44	0,54	3
MA065	B23	5	-0,469	0,006	1,04	1,10	0,46	0,61	3
MA066	B21	4	0,770	0,006	1,08	1,21	0,43	0,39	4
MA067	B04	1	-3,185	0,012	0,99	1,28	0,31	0,87	1
MA068	B23	7	-0,604	0,006	1,05	1,07	0,45	0,63	3
MA069	B04	2	-1,275	0,009	0,79	0,71	0,61	0,62	2
MA070	B30	1	-0,014	0,071	0,85	0,78	0,60	0,53	3
MA071	B08	2	0,601	0,004	1,06	1,10	0,43	0,44	4
MA072	B25	1	-0,478	0,003	0,88	0,83	0,56	0,61	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA073	B22	5	-3,427	0,011	0,94	0,84	0,31	0,94	1
MA074	B06	1	-2,737	0,011	1,00	1,09	0,36	0,83	1
MA075	B22	4	0,103	0,006	1,08	1,12	0,43	0,50	3
MA076	B20	3	-0,276	0,006	1,12	1,22	0,40	0,57	3
MA077	B16	2	-0,327	0,004	1,06	1,05	0,45	0,58	3
MA078	B10	1	-2,081	0,010	0,89	0,83	0,50	0,74	1
MA079	B27	1	-1,934	0,088	0,88	0,80	0,48	0,82	2
MA080	B25	3	-0,646	0,003	1,07	1,10	0,43	0,64	3
MA081	B19	2	-1,526	0,006	0,82	0,72	0,55	0,77	2
MA082	B25	2	0,445	0,003	0,93	0,94	0,54	0,44	4
MA083	B26	1	-1,063	0,003	1,00	1,02	0,46	0,70	2
MA084	B02	2	0,426	0,004	1,07	1,12	0,42	0,48	4
MA085	B12	2	-2,162	0,010	0,94	0,89	0,46	0,76	1
MA086	B04	3	-0,905	0,009	0,98	0,98	0,50	0,55	3
MA087	B10	3	-1,311	0,009	1,14	1,19	0,39	0,62	2
MA088	B07	1	1,122	0,005	1,18	1,47	0,31	0,35	4
MA089	B15	3	-	-	-	-	-	-	-
MA090	B10	2	-1,956	0,009	0,93	0,93	0,48	0,72	2
MA091	B08	1	-0,052	0,004	0,99	0,97	0,48	0,56	3
MA092	B01	3	1,320	0,005	1,22	1,50	0,28	0,32	4
MA093	B05	3	-1,953	0,010	1,05	1,22	0,39	0,73	2
MA094	B05	2	-1,483	0,009	1,08	1,26	0,39	0,66	2
MA095	B11	3	-0,671	0,009	1,02	1,05	0,48	0,52	3
MA096	B26	2	0,156	0,003	1,20	1,31	0,36	0,49	3
MA097	B09	1	0,105	0,004	1,16	1,25	0,37	0,54	3
MA098	B13	3	-0,297	0,004	0,90	0,86	0,55	0,57	3
MA099	B03	3	0,374	0,004	0,96	0,97	0,50	0,49	4
MA100	B03	1	0,314	0,004	1,15	1,22	0,37	0,50	3
MA101	B09	2	1,206	0,005	1,02	1,14	0,44	0,34	4
MA102	B29	1	0,316	0,071	0,83	0,78	0,61	0,47	3
MA103	B28	1	2,003	0,084	0,91	0,77	0,49	0,21	4
MA104	B20	7	-0,376	0,006	0,88	0,84	0,56	0,59	3
MA105	B23	6	-0,233	0,006	0,90	0,89	0,55	0,56	3
MA106	B19	1	-2,596	0,008	1,02	1,10	0,32	0,88	1
MA107	B19	8	-1,628	0,007	0,92	0,82	0,48	0,78	2
MA108	B21	6	-0,673	0,006	0,89	0,82	0,56	0,64	3
MA109	B15	1	-2,367	0,005	1,06	1,26	0,32	0,86	1

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA110	B24	1	-1,360	0,006	0,77	0,66	0,60	0,74	2
MA111	B05	1	-1,476	0,009	1,01	1,05	0,45	0,66	2
MA112	B07	2	1,079	0,005	1,15	1,34	0,34	0,36	4

Los ítems, CT081 de la prueba de Lectura, MA063 y MA089 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 4.º grado de primaria.

Tabla 4.3 Dificultad y ajuste de los ítems al modelo Rasch, Lectura – 6.º grado de primaria

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT001	B02	1	0,110	0,014	1,04	1,09	0,45	0,63	2
CT002	B02	2	-1,326	0,018	0,87	0,62	0,45	0,87	1
CT003	B02	3	-0,230	0,014	1,06	1,17	0,41	0,70	2
CT004	B02	4	-1,276	0,018	1,07	1,46	0,30	0,86	1
CT005	B02	5	-1,080	0,017	0,95	0,95	0,41	0,84	1
CT006	B06	4	1,112	0,013	0,94	0,99	0,54	0,42	4
CT007	B05	5	-0,613	0,015	0,94	0,84	0,50	0,75	2
CT008	B06	5	0,213	0,013	1,11	1,14	0,42	0,61	3
CT009	B07	1	-1,424	0,024	0,79	0,52	0,51	0,87	1
CT010	B07	2	0,437	0,017	1,00	1,00	0,50	0,55	3
CT011	B07	3	-0,858	0,020	1,00	0,90	0,44	0,80	2
CT012	B07	4	-0,784	0,020	0,81	0,67	0,56	0,79	2
CT013	B07	5	0,874	0,017	1,14	1,26	0,41	0,45	3
CT014	B07	6	2,497	0,096	0,83	0,59	0,53	0,16	4
CT015	B11	1	0,291	0,015	1,09	1,12	0,39	0,62	3
CT016	B11	2	1,037	0,015	1,16	1,30	0,36	0,45	4
CT017	B11	3	0,392	0,015	1,05	1,08	0,43	0,60	3
CT018	B11	5	0,094	0,016	0,90	0,83	0,52	0,66	2
CT019	B11	6	0,842	0,015	1,11	1,17	0,40	0,49	3
CT020	B13	1	-0,051	0,016	0,91	0,86	0,48	0,71	2
CT021	B13	2	1,362	0,015	1,07	1,17	0,38	0,39	4
CT022	B13	4	1,114	0,015	1,14	1,20	0,33	0,44	4
CT023	B11	4	0,766	0,015	0,97	0,97	0,50	0,51	3
CT024	B16	1	-0,172	0,018	1,03	1,05	0,48	0,56	2
CT025	B16	2	-0,025	0,018	1,00	0,99	0,50	0,53	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT026	B16	3	-0,395	0,019	0,82	0,74	0,60	0,61	2
CT027	B16	5	0,560	0,019	1,10	1,19	0,44	0,40	3
CT028	B16	6	0,060	0,018	1,13	1,23	0,42	0,51	2
CT029	B01	5	-1,433	0,019	0,85	0,69	0,47	0,87	1
CT030	B10	6	0,666	0,017	1,06	1,07	0,48	0,52	3
CT031	B10	4	0,351	0,017	0,80	0,73	0,62	0,58	3
CT032	B10	5	-0,104	0,018	0,91	0,83	0,53	0,68	2
CT033	B08	5	-0,473	0,018	0,97	0,96	0,44	0,75	2
CT034	B16	4	-0,970	0,020	0,79	0,66	0,58	0,73	1
CT035	B15	1	0,405	0,016	0,94	0,92	0,52	0,61	3
CT036	B15	2	1,112	0,015	1,10	1,17	0,44	0,46	4
CT037	B15	3	0,546	0,015	1,06	1,08	0,45	0,58	3
CT038	B15	4	-0,261	0,017	0,98	0,94	0,45	0,74	2
CT039	B15	5	0,568	0,015	1,02	1,04	0,47	0,58	3
CT040	B15	6	0,652	0,015	1,07	1,13	0,44	0,56	3
CT041	B12	2	-0,416	0,017	0,95	0,85	0,44	0,77	2
CT042	B12	3	1,301	0,015	1,04	1,16	0,44	0,40	4
CT043	B12	4	0,045	0,016	1,02	1,09	0,41	0,69	2
CT044	B14	1	0,994	0,015	1,08	1,13	0,38	0,46	4
CT045	B14	3	0,410	0,015	0,95	0,93	0,48	0,60	3
CT046	B14	5	0,406	0,015	1,10	1,16	0,36	0,60	3
CT047	B05	6	1,071	0,070	0,98	0,95	0,50	0,41	4
CT048	B12	1	-0,286	0,017	0,99	0,95	0,41	0,75	2
CT049	B12	5	0,455	0,015	1,01	0,99	0,45	0,60	3
CT050	B12	6	0,393	0,015	1,02	1,02	0,44	0,61	3
CT051	B18	1	-0,158	0,018	1,02	1,02	0,47	0,61	2
CT052	B18	2	-0,155	0,018	0,97	0,94	0,50	0,60	2
CT053	B18	3	-0,346	0,019	0,88	0,80	0,55	0,64	2
CT054	B18	4	0,154	0,018	1,18	1,22	0,38	0,54	3
CT055	B18	5	-0,631	0,020	0,96	0,93	0,48	0,70	2
CT056	B18	6	0,276	0,018	1,14	1,20	0,41	0,51	3
CT057	B01	4	-1,316	0,018	0,99	1,19	0,38	0,86	1
CT058	B01	3	-0,868	0,016	0,78	0,61	0,57	0,79	2
CT059_1	B01	6	-0,776	0,051	0,90	0,95	0,64	1,29	2
CT059_2	B01	6	0,734	0,051	0,90	0,95	0,64	1,29	3
CT060	B08	1	0,531	0,016	1,03	1,05	0,43	0,53	3
CT061	B08	2	-0,958	0,021	0,84	0,70	0,50	0,83	1

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT062	B08	3	0,563	0,016	1,05	1,06	0,42	0,52	3
CT063	B08	6	0,490	0,070	0,94	0,90	0,51	0,53	3
CT064	B08	4	-0,730	0,019	0,98	1,02	0,41	0,79	2
CT065	B17	3	-1,486	0,023	0,88	0,71	0,48	0,83	1
CT066	B17	4	-0,903	0,020	0,87	0,76	0,53	0,73	1
CT067	B06	1	0,501	0,013	1,08	1,13	0,45	0,55	3
CT068	B06	2	0,203	0,013	0,92	0,88	0,53	0,61	3
CT069	B06	3	0,861	0,013	0,87	0,85	0,59	0,47	3
CT070	B06	6	2,315	0,086	1,08	1,01	0,42	0,19	4
CT071	B10	1	-0,385	0,018	0,82	0,67	0,56	0,73	2
CT072	B10	2	0,776	0,017	0,95	0,97	0,54	0,49	3
CT073	B10	3	-0,204	0,018	0,97	0,95	0,48	0,70	2
CT074	B05	1	0,119	0,013	0,89	0,85	0,56	0,61	2
CT075	B05	2	0,244	0,013	1,01	1,02	0,48	0,58	3
CT076	B05	3	0,016	0,014	1,26	1,41	0,32	0,63	2
CT077	B05	4	0,807	0,013	1,12	1,25	0,42	0,46	3
CT078	B09	1	-1,404	0,022	0,98	0,97	0,41	0,85	1
CT079	B09	2	-0,219	0,017	1,01	0,98	0,49	0,65	2
CT080	B09	4	-0,227	0,017	0,91	0,84	0,55	0,65	2
CT081	B09	5	0,327	0,017	1,17	1,29	0,40	0,54	3
CT082	B04	4	0,182	0,016	0,95	0,94	0,48	0,61	3
CT083	B04	5	1,521	0,017	1,10	1,26	0,34	0,30	4
CT084	B04	6	0,314	0,070	0,94	0,96	0,50	0,57	3
CT085	B17	1	-0,673	0,019	0,98	0,99	0,47	0,69	2
CT086	B03	1	-0,469	0,018	0,93	0,84	0,51	0,73	2
CT087	B03	2	-	-	-	-	-	-	-
CT088	B03	4	-0,146	0,017	0,96	0,90	0,51	0,67	2
CT089	B03	5	0,673	0,017	1,08	1,15	0,45	0,50	3
CT090	B13	3	1,006	0,015	1,02	1,06	0,42	0,47	4
CT091	B14	2	0,565	0,015	1,04	1,06	0,41	0,56	3
CT092	B01	1	-1,254	0,018	0,86	0,73	0,48	0,85	1
CT093	B09	3	-0,133	0,017	1,00	0,99	0,50	0,63	2
CT094	B09	6	0,046	0,017	1,00	1,01	0,50	0,60	2
CT095	B14	4	1,413	0,015	1,09	1,24	0,36	0,36	4
CT096	B14	6	0,346	0,015	0,97	0,96	0,46	0,61	3
CT097	B03	3	-1,277	0,023	0,89	0,72	0,47	0,86	1
CT098	B01	2	-0,928	0,016	0,92	0,85	0,48	0,80	1

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
CT099	B02	6	-0,073	0,072	0,89	0,82	0,53	0,67	2
CT100	B03	6	-0,427	0,080	0,90	0,96	0,52	0,72	2
CT101	B04	1	-0,235	0,017	0,94	0,91	0,47	0,70	2
CT102	B04	2	1,209	0,017	1,06	1,21	0,38	0,37	4
CT103	B04	3	-0,92	0,020	0,88	0,73	0,48	0,82	1
CT104	B13	5	0,082	0,016	0,96	0,93	0,45	0,68	2
CT105	B13	6	0,978	0,015	1,09	1,16	0,37	0,48	4
CT106	B17	2	-0,292	0,019	0,90	0,86	0,55	0,61	2
CT107	B17	5	0,651	0,019	1,01	1,08	0,50	0,41	3
CT108	B17	6	0,018	0,018	1,02	1,04	0,48	0,54	2

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	B09	2	-1,084	0,020	1,02	1,02	0,46	0,73	1
MA002	B10	2	-1,603	0,022	0,84	0,74	0,51	0,81	1
MA003	B20	10	-1,461	0,017	0,93	0,80	0,42	0,85	1
MA004	B07	2	-1,944	0,024	0,87	0,71	0,45	0,86	1
MA005	B08	1	-1,137	0,020	0,98	0,91	0,46	0,75	1
MA006	B08	3	-1,235	0,021	0,82	0,66	0,55	0,76	1
MA007	B19	3	-0,527	0,014	1,01	1,03	0,40	0,74	2
MA008	B19	8	-0,452	0,014	0,96	1,01	0,43	0,72	2
MA009	B22	2	-0,899	0,019	0,88	0,84	0,50	0,77	2
MA010	B22	8	-0,938	0,019	0,94	0,89	0,47	0,78	1
MA011	B19	7	0,318	0,013	0,88	0,85	0,53	0,58	2
MA012	B21	2	-0,753	0,019	0,91	0,86	0,50	0,75	2
MA013	B22	3	0,94	0,017	1,18	1,29	0,39	0,44	3
MA014	B19	9	-1,284	0,017	0,92	0,92	0,40	0,84	1
MA015	B28	1	1,663	0,016	0,98	1,01	0,50	0,35	4
MA016	B20	8	-0,543	0,014	0,96	1,04	0,46	0,73	2
MA017	B20	1	-1,158	0,016	0,96	1,04	0,41	0,82	1
MA018	B19	4	0,537	0,013	1,03	1,05	0,43	0,54	3
MA019	B21	4	1,406	0,017	1,08	1,19	0,44	0,35	4
MA020	B26	2	-0,639	0,017	0,91	0,88	0,44	0,78	2
MA021	B20	3	-0,431	0,014	1,00	1,21	0,43	0,71	2
MA022	B21	7	0,721	0,017	1,04	1,09	0,47	0,48	3
MA023	B20	2	-0,634	0,014	0,97	0,92	0,46	0,74	2

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA024	B20	6	1,08	0,013	0,99	1,03	0,50	0,42	3
MA025	B20	5	1,086	0,013	1,07	1,15	0,45	0,42	3
MA026	B22	1	-1,444	0,021	0,80	0,62	0,51	0,84	1
MA027	B22	7	-0,832	0,019	0,90	0,88	0,50	0,76	2
MA028	B07	3	-1,465	0,021	0,88	0,74	0,48	0,80	1
MA029	B04	3	0,489	0,015	0,97	0,95	0,48	0,58	3
MA030	B07	1	-0,206	0,018	1,04	1,02	0,46	0,59	2
MA031	B32	1	-2,060	0,025	0,87	0,76	0,44	0,86	1
MA032	B32	2	-1,269	0,021	0,96	0,90	0,46	0,77	1
MA033	B06	2	1,896	0,016	0,88	0,90	0,53	0,30	4
MA034	B03	3	1,443	0,015	1,04	1,14	0,46	0,39	4
MA035	B31	1	-0,796	0,019	1,19	1,47	0,31	0,69	2
MA036	B09	3	-0,791	0,020	0,87	0,76	0,57	0,69	2
MA037	B22	10	-0,798	0,019	0,98	0,91	0,46	0,76	2
MA038	B26	1	1,214	0,015	0,99	1,02	0,47	0,43	3
MA039	B20	7	0,845	0,013	1,03	1,09	0,47	0,47	3
MA040	B25	2	0,670	0,015	1,15	1,25	0,32	0,53	3
MA041	B09	1	-1,599	0,022	0,98	0,99	0,44	0,80	1
MA042	B21	8	-0,574	0,018	0,89	0,81	0,53	0,72	2
MA043	B04	1	0,656	0,015	0,99	0,99	0,47	0,55	3
MA044	B34	1	-1,649	0,022	1,08	1,31	0,36	0,81	1
MA045	B22	6	0,663	0,016	1,03	1,09	0,48	0,49	3
MA046	B21	10	-0,950	0,019	0,96	0,92	0,46	0,78	1
MA047	B10	1	-0,973	0,020	0,86	0,72	0,56	0,72	1
MA048	B01	2	0,604	0,015	0,95	0,93	0,47	0,55	3
MA049	B19	10	0,320	0,013	1,02	1,02	0,44	0,58	2
MA050	B33	2	-1,728	0,023	0,88	0,86	0,48	0,82	1
MA051	B31	2	-1,088	0,020	1,07	1,03	0,40	0,74	1
MA052	B21	1	-1,159	0,020	0,80	0,70	0,54	0,80	1
MA053	B19	1	-1,107	0,016	0,90	0,82	0,43	0,82	1
MA054	B20	4	0,809	0,013	0,92	0,94	0,54	0,48	3
MA055	B28	2	0,223	0,015	0,89	0,83	0,52	0,63	2
MA056	B21	9	-0,308	0,017	0,99	0,96	0,48	0,67	2
MA057	B33	1	-1,592	0,022	0,97	1,01	0,44	0,80	1
MA058	B06	3	0,648	0,015	0,99	0,98	0,46	0,54	3
MA059	B34	2	-1,465	0,022	0,88	0,76	0,50	0,79	1
MA060	B29	2	1,050	0,015	1,04	1,06	0,43	0,47	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA061	B22	4	0,782	0,016	0,97	1,00	0,51	0,47	3
MA062	B20	9	-0,012	0,013	0,97	0,95	0,49	0,63	2
MA063_1	B15	1	0,840	0,045	0,88	0,81	0,68	0,78	3
MA063_2	B15	1	1,965	0,045	0,88	0,81	0,68	0,78	4
MA064	B30	2	1,437	0,015	0,95	0,96	0,49	0,38	4
MA065	B06	1	-0,280	0,016	1,00	1,02	0,40	0,72	2
MA066	B22	9	-	-	-	-	-	-	-
MA067	B03	1	-	-	-	-	-	-	-
MA068	B14	1	1,493	0,072	0,86	0,74	0,59	0,35	4
MA069	B05	3	2,044	0,017	0,95	1,06	0,48	0,28	4
MA070	B17	1	-0,227	0,077	0,80	0,71	0,61	0,59	2
MA071	B03	2	0,410	0,015	1,01	1,02	0,46	0,59	3
MA072	B08	2	0,907	0,019	1,01	1,09	0,50	0,38	3
MA073	B27	2	0,492	0,015	0,84	0,78	0,57	0,58	3
MA074	B27	1	0,232	0,015	0,94	0,90	0,50	0,63	2
MA075	B30	1	1,578	0,016	1,00	1,03	0,46	0,35	4
MA076_1	B18	1	-0,817	0,054	0,94	0,94	0,68	1,15	2
MA076_2	B18	1	0,426	0,054	0,94	0,94	0,68	1,15	3
MA077	B21	6	0,906	0,017	1,05	1,17	0,46	0,45	3
MA078	B01	1	1,063	0,015	1,15	1,21	0,33	0,45	3
MA079	B05	2	1,670	0,016	1,28	1,49	0,26	0,35	4
MA080	B21	3	0,319	0,017	1,09	1,17	0,43	0,56	2
MA081	B10	3	-0,754	0,019	0,99	0,94	0,49	0,68	2
MA082	B02	3	1,500	0,015	0,89	0,86	0,54	0,38	4
MA083	B16	1	0,859	0,068	0,85	0,88	0,58	0,43	3
MA084	B25	1	2,402	0,018	1,06	1,31	0,37	0,22	4
MA085	B04	2	1,894	0,016	1,04	1,16	0,44	0,31	4
MA086	B02	2	1,415	0,015	1,08	1,18	0,40	0,39	4
MA087	B19	2	-0,545	0,014	0,97	0,98	0,42	0,74	2
MA088	B01	3	0,676	0,015	1,08	1,10	0,38	0,53	3
MA089	B05	1	1,147	0,015	0,99	1,02	0,46	0,45	3
MA090	B22	5	1,160	0,017	1,14	1,28	0,41	0,40	3
MA091	B19	6	1,155	0,013	1,11	1,28	0,38	0,41	3
MA092	B24	1	1,036	0,014	1,04	1,22	0,48	0,35	3
MA093	B24	2	0,483	0,013	1,00	1,03	0,51	0,45	3
MA094	B24	3	0,965	0,013	0,99	1,11	0,51	0,37	3
MA095	B23	1	0,731	0,009	1,01	1,02	0,45	0,53	3

Ítem	Bloque	Posición	Medida	Error	Infit	Outfit	PTME	p	Nivel
MA096	B23	2	1,308	0,009	0,99	1,01	0,48	0,41	3
MA097	B23	3	1,885	0,009	1,15	1,39	0,34	0,31	4
MA098	B19	5	1,233	0,013	1,11	1,24	0,38	0,40	3
MA099	B21	5	1,249	0,017	0,95	0,99	0,52	0,38	3
MA100	B29	1	-0,418	0,017	0,91	0,85	0,46	0,75	2
MA101_1	B12	1	0,744	0,060	1,01	1,00	0,52	0,56	3
MA101_2	B12	1	4,870	0,060	1,01	1,00	0,52	0,56	4
MA102_1	B13	1	0,854	0,053	1,11	1,11	0,53	0,62	3
MA102_2	B13	1	3,211	0,053	1,11	1,11	0,53	0,62	4
MA103	B11	1	2,768	0,084	0,76	0,62	0,59	0,17	4
MA104	B02	1	2,104	0,017	0,98	0,98	0,48	0,27	4

Los ítems, CT087 de la prueba de Lectura, MA066 y MA067 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 6.º grado de primaria.

4.3.2 Evidencias de validez vinculadas a la estructura interna

Para la ENLA 2024, 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 4.5.

Tabla 4.5 Análisis de unidimensionalidad de las medidas derivadas de las pruebas de la ENLA 2024

Prueba	Varianza de la dimensión principal (%)	Varianza modelada (%)	Primer autovalor	Primer autovalor (%)
Lectura, 4.º grado de primaria	26,7	26,6	1,83	1,4
Matemática, 4.º grado de primaria	31,9	31,7	1,82	1,1
Lectura, 6.º grado de primaria	28,6	28,7	1,44	1,0
Matemática, 6.º grado de primaria	32,4	32,1	1,40	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 4.6 Análisis de confiabilidad de las medidas derivadas de la aplicación de las pruebas de la ENLA 2024

Prueba	Rp	Gp
Lectura, 4.° grado de primaria	0,90	3,02
Matemática, 4.° grado de primaria	0,88	2,74
Lectura, 6.° grado de primaria	0,80	1,98
Matemática, 6.° grado de primaria	0,82	2,12

Considerando el valor de Rp para la aplicación de la ENLA 2024, se observa que la varianza de error se encuentra entre 10% y 20% para las pruebas aplicadas. 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 a continuación.

Tabla 4.7 Indicadores del análisis de precisión y consistencia de la clasificación usando las medidas derivadas de las pruebas de 4.° grado de primaria

	Lectura		Matemática	
	Precisión	Consistencia	Precisión	Consistencia
Nacional	0,79	0,71	0,81	0,74
Hombre	0,79	0,71	0,82	0,75
Mujer	0,79	0,71	0,81	0,73
Urbano	0,79	0,71	0,81	0,74
Rural	0,77	0,68	0,81	0,74
Estatal	0,79	0,70	0,81	0,74
No estatal	0,80	0,72	0,82	0,74
Polidocente Completo	0,79	0,71	0,81	0,74
Unidocente / Multigrado	0,77	0,69	0,81	0,74
Alta habilidad	0,79	0,71	0,81	0,74
Baja habilidad	0,77	0,68	0,81	0,74

Tabla 4.8 Indicadores del análisis de precisión y consistencia de la clasificación usando las medidas derivadas de las pruebas de 6.º grado de primaria

	Lectura		Matemática	
	Precisión	Consistencia	Precisión	Consistencia
Nacional	0,66	0,56	0,72	0,61
Hombre	0,67	0,56	0,72	0,62
Mujer	0,66	0,55	0,61	0,71
Urbano	0,66	0,55	0,71	0,61
Rural	0,68	0,58	0,72	0,63
Estatal	0,66	0,56	0,72	0,61
No estatal	0,66	0,56	0,71	0,60
Polidocente Completo	0,66	0,55	0,71	0,61
Unidocente / Multigrado	0,69	0,59	0,74	0,64
Alta habilidad	0,65	0,55	0,71	0,60
Baja habilidad	0,68	0,58	0,73	0,63

4.4 Equiparación de medidas

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 en la primera mitad de los años ochenta) 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; y Zhu, 1998), se puede señalar que existen tres tipos de diseño 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 4.º grado y 6.º grado de primaria en la misma métrica que el de las pruebas aplicadas en el 2016 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 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:

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

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, 2014). 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.

Las tablas, desde 4.9 hasta 4.13, presentan los resultados del proceso de equiparación por ítems comunes de las pruebas aplicadas en la ENLA 2024. Dichos resultados nos permiten señalar que este proceso ha sido realizado siguiendo rigurosos criterios técnicos, que sustentan la comparación entre los resultados de las evaluaciones realizadas en años diferentes.

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

Tabla 4.9 Análisis de FDI de las medidas, prueba de Lectura de 4.° grado de primaria, usadas en la equiparación de 2023 con 2024

Ítem	2023		2024		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT001	1,023	0,012	1,302	0,006	0,896	0,127	0,013
CT002	0,615	0,012	1,224	0,006	0,825	-0,210	0,013
CT003	-0,543	0,011	0,008	0,005	-0,289	-0,254	0,012
CT004	-1,706	0,013	-1,272	0,007	-1,462	-0,244	0,015
CT009	-0,807	0,011	-0,534	0,006	-0,786	-0,021	0,013
CT027	-1,134	0,007	-0,506	0,003	-0,760	-0,374	0,008
CT028	-1,223	0,01	-0,897	0,004	-1,119	-0,104	0,011
CT030	0,226	0,011	0,481	0,005	0,144	0,082	0,012
CT031	-0,998	0,012	-0,814	0,006	-1,042	0,044	0,013
CT032	-1,408	0,012	-1,128	0,006	-1,330	-0,078	0,013
CT033	-0,936	0,011	-0,712	0,006	-0,949	0,013	0,013
CT034	-0,416	0,011	-0,274	0,006	-0,548	0,132	0,013
CT035	-1,356	0,012	-1,261	0,007	-1,452	0,096	0,014
CT036	-1,004	0,012	-0,690	0,006	-0,929	-0,075	0,013
CT037	-0,264	0,011	-0,008	0,006	-0,304	0,040	0,013
CT038	-0,664	0,011	-0,483	0,006	-0,739	0,075	0,013
CT039	0,283	0,011	0,835	0,005	0,468	-0,185	0,012
CT040	0,184	0,011	0,615	0,005	0,267	-0,083	0,012
CT041	-1,294	0,015	-1,226	0,007	-1,420	0,126	0,017
CT042	-0,796	0,014	-0,720	0,006	-0,956	0,160	0,015
CT043	-1,862	0,017	-1,882	0,008	-2,021	0,159	0,019
CT044	-1,966	0,018	-2,055	0,008	-2,179	0,213	0,020
CT061	0,646	0,013	0,729	0,005	0,371	0,275	0,014
CT062	-1,608	0,016	-1,591	0,007	-1,754	0,146	0,017
CT073	0,566	0,011	0,864	0,005	0,495	0,071	0,012
CT087	-1,607	0,007	-1,110	0,004	-1,314	-0,293	0,008
CT088	-0,404	0,006	-0,012	0,003	-0,308	-0,096	0,007
CT089_1	0,138	0,058	0,206	0,050	-0,108	0,246	0,077
CT089_2	1,277	0,058	1,858	0,050	1,405	-0,128	0,077
CT090	0,416	0,009	0,462	0,003	0,126	0,290	0,009
CT091	-1,096	0,010	-1,176	0,004	-1,374	0,278	0,011
CT092	0,740	0,010	0,900	0,003	0,528	0,212	0,010
CT094	-1,268	0,011	-1,146	0,004	-1,347	0,079	0,012

Ítem	2023		2024		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT096	0,384	0,006	0,835	0,003	0,468	-0,084	0,007
CT098	-0,130	0,006	0,306	0,003	-0,016	-0,114	0,007
CT099	-1,046	0,007	-0,521	0,003	-0,774	-0,272	0,008
CT100	-0,554	0,007	-0,008	0,003	-0,304	-0,250	0,008

Tabla 4.10 Análisis de FDI de las medidas, prueba de Matemática de 4.º grado de primaria, usadas en la equiparación de 2023 con 2024

Ítem	2023		2024		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA004	-1,582	0,013	-0,724	0,006	-1,466	-0,116	0,014
MA008	-2,051	0,013	-1,306	0,007	-2,078	0,027	0,015
MA027	-1,701	0,013	-0,997	0,007	-1,753	0,052	0,015
MA031	-1,550	0,012	-0,524	0,004	-1,256	-0,294	0,013
MA035	-2,130	0,014	-1,432	0,007	-2,210	0,080	0,016
MA036	-1,907	0,013	-1,397	0,007	-2,173	0,266	0,015
MA040	-0,788	0,012	-0,297	0,006	-1,017	0,229	0,013
MA041	0,153	0,011	0,681	0,006	0,010	0,143	0,013
MA042	-2,771	0,016	-1,697	0,008	-2,488	-0,283	0,018
MA044	0,556	0,012	1,012	0,006	0,358	0,198	0,013
MA046	1,201	0,013	1,935	0,006	1,328	-0,127	0,014
MA047	0,397	0,012	1,058	0,004	0,406	-0,009	0,013
MA048	0,641	0,012	1,335	0,006	0,697	-0,056	0,013
MA050	-0,653	0,009	0,061	0,006	-0,641	-0,012	0,011
MA053	0,713	0,012	1,435	0,006	0,802	-0,089	0,013
MA054	-1,086	0,012	-0,621	0,006	-1,358	0,272	0,013
MA055	0,072	0,009	0,904	0,006	0,245	-0,173	0,011
MA058	-0,417	0,009	0,110	0,006	-0,590	0,173	0,011
MA059	-0,283	0,011	0,443	0,003	-0,240	-0,043	0,011
MA060	-0,116	0,011	0,768	0,006	0,102	-0,218	0,013
MA061	-1,292	0,012	-0,779	0,006	-1,524	0,232	0,013
MA065	-0,753	0,012	0,225	0,006	-0,469	-0,284	0,013
MA066	0,930	0,012	1,403	0,006	0,769	0,161	0,013
MA068	-0,882	0,009	0,096	0,006	-0,604	-0,278	0,011
MA073	-3,154	0,013	-2,590	0,011	-3,427	0,273	0,017

Ítem	2023		2024		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA075	0,447	0,012	0,769	0,006	0,103	0,344	0,013
MA076	-0,248	0,009	0,408	0,006	-0,277	0,029	0,011
MA081	-1,253	0,009	-0,781	0,006	-1,526	0,273	0,011
MA104	-0,660	0,012	0,313	0,006	-0,376	-0,284	0,013
MA105	-0,241	0,011	0,449	0,006	-0,234	-0,007	0,013
MA106	-2,982	0,012	-1,799	0,008	-2,596	-0,386	0,014
MA107	-1,949	0,013	-0,878	0,007	-1,628	-0,321	0,015
MA108	-0,702	0,012	0,030	0,006	-0,674	-0,028	0,013
MA109	-2,208	0,014	-1,581	0,005	-2,367	0,159	0,015
MA110	-1,262	0,012	-0,623	0,006	-1,360	0,098	0,013

Tabla 4.11 Análisis de FDI de las medidas, prueba de Lectura de 6.º grado de primaria, usadas en la equiparación de 2022 con 2024

Ítem	2022		2024		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT001	0,361	0,050	0,028	0,014	0,110	0,251	0,052
CT003	0,190	0,051	-0,382	0,014	-0,230	0,420	0,053
CT004	-0,874	0,063	-1,644	0,018	-1,276	0,402	0,066
CT005	-0,795	0,062	-1,408	0,017	-1,080	0,285	0,064
CT006	0,800	0,056	1,236	0,013	1,111	-0,311	0,057
CT007	-0,972	0,072	-0,844	0,015	-0,613	-0,359	0,074
CT008	0,253	0,057	0,152	0,013	0,213	0,040	0,058
CT009	-1,684	0,083	-1,823	0,024	-1,424	-0,260	0,086
CT010	0,380	0,044	0,422	0,017	0,436	-0,056	0,047
CT011	-1,060	0,064	-1,140	0,020	-0,858	-0,202	0,067
CT012	-0,849	0,059	-1,051	0,020	-0,784	-0,065	0,062
CT013	1,108	0,043	0,949	0,017	0,873	0,235	0,046
CT029	-1,395	0,068	-1,833	0,019	-1,433	0,038	0,071
CT033	-0,689	0,071	-0,676	0,018	-0,474	-0,215	0,073
CT057	-1,130	0,062	-1,692	0,018	-1,316	0,186	0,065
CT058	-0,901	0,057	-1,152	0,016	-0,868	-0,033	0,059
CT059_1	-0,950	0,052	-0,865	0,051	-0,630	-0,320	0,073
CT059_2	0,714	0,052	0,604	0,051	0,587	0,127	0,073
CT060	0,559	0,056	0,535	0,016	0,530	0,029	0,058

Ítem	2022		2024		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
CT061	-1,095	0,081	-1,260	0,021	-0,958	-0,137	0,084
CT062	0,796	0,056	0,574	0,016	0,562	0,234	0,058
CT063	0,537	0,112	0,486	0,070	0,489	0,048	0,132
CT064	-0,934	0,077	-0,985	0,019	-0,730	-0,204	0,079
CT067	0,396	0,056	0,499	0,013	0,500	-0,104	0,057
CT068	0,144	0,058	0,140	0,013	0,203	-0,059	0,059
CT069	0,702	0,056	0,934	0,013	0,861	-0,159	0,057
CT070	1,921	0,100	2,687	0,086	2,314	-0,393	0,132
CT074	0,157	0,057	0,039	0,013	0,119	0,038	0,058
CT075	0,017	0,058	0,189	0,013	0,243	-0,226	0,059
CT076	0,002	0,058	-0,086	0,014	0,015	-0,013	0,060
CT077	0,852	0,055	0,869	0,013	0,807	0,045	0,057
CT086	-0,216	0,063	-0,671	0,018	-0,469	0,253	0,066
CT088	0,167	0,060	-0,281	0,017	-0,146	0,313	0,062
CT089	0,858	0,057	0,707	0,017	0,673	0,185	0,059
CT092	-1,263	0,065	-1,618	0,018	-1,254	-0,009	0,067
CT098	-0,931	0,058	-1,224	0,016	-0,928	-0,003	0,060

Tabla 4.12 Análisis de FDI de las medidas, prueba de Matemática de 6.º grado de primaria, usadas en la equiparación de 2022 con 2024

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA007	-0,234	0,046	-0,774	0,014	-0,527	0,293	0,048
MA008	-0,531	0,048	-0,696	0,014	-0,453	-0,078	0,050
MA009	-0,978	0,066	-1,163	0,019	-0,900	-0,078	0,069
MA010	-1,053	0,054	-1,204	0,019	-0,939	-0,114	0,057
MA011	0,304	0,055	0,109	0,013	0,318	-0,014	0,057
MA012	-0,918	0,053	-1,010	0,019	-0,753	-0,165	0,056
MA013	0,688	0,041	0,759	0,017	0,94	-0,252	0,044
MA014	-1,176	0,069	-1,565	0,017	-1,285	0,109	0,071
MA016	-0,407	0,047	-0,791	0,014	-0,544	0,137	0,049
MA017	-1,049	0,070	-1,434	0,016	-1,159	0,110	0,072
MA018	0,745	0,055	0,337	0,013	0,536	0,209	0,057
MA019	1,401	0,056	1,246	0,017	1,406	-0,005	0,059

Ítem	2022		2023		EQP	FDI	
	Medida	Error	Medida	Error	Medida equiparada	Medida	Error
MA021	-0,178	0,045	-0,674	0,014	-0,432	0,254	0,047
MA022	0,590	0,041	0,530	0,017	0,721	-0,131	0,044
MA023	-0,491	0,060	-0,886	0,014	-0,635	0,144	0,062
MA024	0,992	0,041	0,905	0,013	1,080	-0,088	0,043
MA025	1,347	0,056	0,911	0,013	1,086	0,261	0,057
MA026	-1,538	0,079	-1,732	0,021	-1,445	-0,093	0,082
MA027	-0,578	0,049	-1,093	0,019	-0,833	0,255	0,053
MA037	-0,866	0,076	-1,057	0,019	-0,798	-0,068	0,078
MA039	0,694	0,039	0,659	0,013	0,844	-0,150	0,041
MA042	-0,786	0,046	-0,823	0,018	-0,574	-0,212	0,049
MA045	0,758	0,039	0,469	0,016	0,663	0,095	0,042
MA046	-0,838	0,075	-1,216	0,019	-0,951	0,113	0,077
MA052	-1,399	0,060	-1,435	0,020	-1,160	-0,239	0,063
MA053	-1,146	0,071	-1,380	0,016	-1,108	-0,038	0,073
MA054	0,428	0,039	0,622	0,013	0,809	-0,381	0,041
MA077	0,812	0,039	0,723	0,017	0,906	-0,094	0,043
MA080	0,463	0,039	0,110	0,017	0,319	0,144	0,043
MA087	-0,514	0,043	-0,793	0,014	-0,546	0,032	0,045
MA090	1,325	0,056	0,989	0,017	1,160	0,165	0,059
MA091	1,061	0,055	0,983	0,013	1,155	-0,094	0,057
MA098	1,365	0,043	1,065	0,013	1,233	0,132	0,045
MA099	1,092	0,039	1,082	0,017	1,249	-0,157	0,043

Tabla 4.13 Constantes de equiparación de las pruebas de la ENLA 2024

Prueba	a	b
Lectura, 4.° grado de primaria	0,916	-0,297
Matemática, 4.° grado de primaria	1,051	-0,705
Lectura, 6.° grado de secundaria	0,829	0,087
Matemática, 6.° grado de secundaria	0,957	0,214

4.5 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) que 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 sostener los mismos puntos de corte en las ediciones posteriores y asegurar que los resultados sean comparables en el tiempo.

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 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 llegar 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. La primera 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. La segunda 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 fue mencionado con anterioridad, 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 del Perú, 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 4.14, se presentan los puntos de corte (en M500) de las evaluaciones ENLA 2024.

Tabla 4.14 Medidas (en M500) que marcan los puntos de corte de los estudiantes en la ENLA 2024

Área	< Nivel 1 vs, Nivel 1	Nivel 1 vs, Nivel 2	Nivel 2 vs Nivel 3
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, 6.º grado de primaria	445,46	522,17	579,14
Matemática, 6.º grado de primaria	432,32	526,61	603,41

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 4.15.

Tabla 4.15 Valores de a y b utilizados para la transformación lineal de las medidas de la ENLA 2024

Área	a	b
Lectura, 4.º grado de primaria	474,896	73,414
Matemática, 4.º grado de primaria	463,157	73,781
Lectura, 6.º grado de primaria	474,896	73,414
Matemática, 6.º grado de primaria	463,157	73,414

Procesamiento de datos de Factores Asociados

Capítulo 5

Capítulo 5

Procesamiento de datos y elaboración de cuestionarios de Factores Asociados

Los factores asociados son un conjunto de variables que contribuyen a comprender y contextualizar los resultados de aprendizaje. 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 como las que realiza PISA, ERCE o la misma ENLA.

En la ENLA 2024, 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 (4.º y 6.º grados de primaria).

Cada uno de los cuestionarios aplicados está compuesto por 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) y que, en algunos casos, pueden ser construidas a partir de la recodificación de una o más preguntas (por ejemplo, años trabajando en la escuela). Por otro lado, también se consideran preguntas que están diseñadas para representar un constructo latente que no puede observarse directamente (por ejemplo: sentido de pertenencia). Estas preguntas están conceptualmente relacionadas entre sí y se agrupan en una única medida compuesta que nos permite representar la variable latente a estudiar (Bandalos, 2018).

Este capítulo busca brindar una descripción de las variables recogidas en el presente año, así como los procedimientos realizados para su elaboración y los aspectos relacionados con el análisis de dicha información.

5.1 Elaboración de las preguntas del cuestionario de factores asociados

Los cuestionarios son un conjunto de instrumentos que permiten 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 propósito de asegurar la calidad del recojo de la información, es importante considerar su correcta elaboración y validación. En ese sentido, la presente sección detalla el proceso realizado para la preparación de los cuestionarios, considerando las cuatro fases señaladas por Furr (2013):

- 1) articulación del constructo y el contexto.
- 2) elaboración de ítems y escalas de respuesta.
- 3) obtención de los datos.
- 4) análisis psicométrico.

En la presente sección se desarrollarán los puntos principales de las dos primeras fases.

1. 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 2024 se quiso realizar un énfasis en las variables que se asocian con los logros de aprendizaje en el área rural, lo que implicó la elaboración de cuestionarios al director diferenciados por área. 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 violencia escolar dado el aumento de casos), por lo que los cuestionarios de factores asociados también deben responder a estas necesidades de información.

Con la definición de los ejes temáticos de los factores asociados en la ENLA 2024, 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 de las IE con la UGEL, aspectos referidos a la implementación del currículo traducidos en prácticas pedagógicas, creencias docentes, oportunidades de aprendizaje, las preocupaciones de los estudiantes por terminar la primaria, el clima escolar y la presencia de violencia en la escuela 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: dotación de equipos tecnológicos en 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 2024 se emplearon las siguientes categorías: a) características y contexto de la escuela; b) creencias y prácticas pedagógicas del docente; c) antecedentes del estudiante y su hogar; y d) motivación, creencias y emociones del estudiante.

Tanto la revisión bibliográfica como la organización de la propuesta fueron los primeros pasos para alcanzar una 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: 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, 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.

2. Elaboración de ítems y escalas de respuesta.

Para la redacción de los ítems, se tomaron en cuenta los criterios de Stehr-Green et al. (2003) y de Maholtra (2016), quienes señalan que estos deben incluir una idea clara y sin ambigüedad. Asimismo, se usó un vocabulario simple y se evitó el lenguaje técnico, las abreviaturas, las negaciones (en la medida de lo posible) 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 ideal. Más allá de esto, el número máximo de ítems por escala dependerá de la complejidad de la variable que se 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 reducir las tasas de omisión por cansancio (Furr, 2013; Robinson, 2018).

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

Una vez elaborada la propuesta final de escalas e ítems, se llevó a cabo un proceso de validación por parte de otros especialistas del Minedu, con el objetivo de verificar su pertinencia. Los especialistas fueron seleccionados según el ámbito educativo de acción; por ejemplo: para el caso de los ítems que recogieron información pedagógica, se buscó 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; por ejemplo, hacer coincidir una respuesta generada internamente con una respuesta de la encuesta (Willis, 2011).

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

4. 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

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.

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, el cual, consiste principalmente en la elaboración de reportes descriptivos, así como en el recojo de evidencias de validez y confiabilidad de las escalas de factores asociados.

5.2 Elaboración de análisis descriptivo

Se calcularon los porcentajes de respuesta en cada una de las opciones de marcado, de cada uno de los ítems incluidos en el cuestionario. Después, las distribuciones de respuesta fueron representadas en gráficos de barras, lo cual fue útil para identificar los 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.

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

Luego de la generación 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 para producir las variables de factores asociados dependen del enfoque que se ha utilizado para conceptualizarlas. Así, se distinguen medidas reflexivas y formativas (Bollen y Bauldry, 2011; Bollen y Diamantopoulos, 2017; Jarvis et al., 2003). En caso de que las respuestas a los ítems representen el efecto de la variable latente, se está frente a un modelo reflexivo. Un ejemplo de este tipo de variables es la escala de ansiedad ante 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

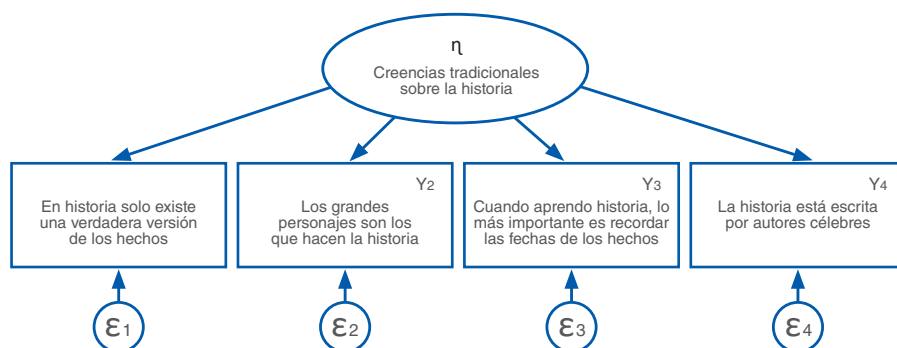
Tal como se señaló anteriormente, las evidencias de validez para constructos reflexivos son recolectadas a través del AFC. Al utilizarlo, se necesita testear el modelo teórico de la dimensionalidad de los ítems por medio de datos empíricos. Este modelo vincula las variables observables (ítems) con la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

² 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 el AFC. 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.

Figura 5.2. Representación gráfica de un constructo formativo

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 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. Debido a la naturaleza de las respuestas de los cuestionarios, se ajustan modelos pertinentes para ítems categóricos. Por ello, el estimador utilizado fue 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 entre los ítems que componen cada indicador. El tratamiento de valores perdidos se realizó mediante eliminación pairwise, siguiendo el procedimiento implementado por el paquete lavaan. Este enfoque permite estimar los parámetros del modelo a partir de información disponible en cada par de variables, sin requerir la imputación de valores faltantes.

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 0,95 de CFI y TLI señalarían un buen ajuste en el modelo (Hu y Bentler, 1999; Marsh et al., 2004; Lai y Green, 2016). De otro lado, valores inferiores a 0,10 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 (R Core Team). 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
4.º y 6.º grado de primaria	Director	DIRP_COLABDOC	Colaboración entre docentes	0,995	0,992	0,035
		DIRPEN_SITPROBR	Situaciones problemáticas en la escuela	0,989	0,988	0,068
		DIRP_AUTIE	Autonomía para realizar mejoras en la escuela	0,996	0,995	0,036
		DIRP_SATUGEL	Satisfacción con el apoyo recibido por la UGEL	0,998	0,997	0,033
		DIRP_LIDUG	Liderazgo educativo de la UGEL	0,997	0,996	0,036
		DIRP_ENSHSE	Creencias sobre la enseñanza de las HSE	0,991	0,986	0,047
		DIRP_CREAUTO	Creencias sobre la autorregulación	0,996	0,991	0,050
		DIRP_UTILLEC	Percepción de utilidad de materiales educativos de Comunicación	0,998	0,993	0,024
		DIRP_UTILMAT	Percepción de utilidad de materiales educativos de Matemática	0,999	0,997	0,019

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR
4.° grado de primaria	Docente de Comunicación	DOC4PLEC_CONFDD	Confianza director-docente	1,000	0,999	0,013
		DOC4PLEC_CRERLEC	Creencias sobre la recursividad de la escritura	0,995	0,991	0,025
		DOC4PLEC_EVFORM	Creencias sobre la evaluación formativa	0,982	0,978	0,050
		DOC4PLEC_IEAPRE	Escuela orientada al desarrollo de aprendizajes	0,998	0,996	0,025
		DOC4PLEC_PRACLECT	Prácticas docentes en la enseñanza de la lectura	0,993	0,991	0,060
		DOC4PLEC_PREPCOM	Preparación del docente para la enseñanza de Lectura	0,997	0,996	0,034
		DOC4PLEC_PROBIE	Percepción de situaciones problemáticas en la escuela	0,990	0,983	0,052
	Docente de Matemática	DOC4PLEC_PROMAU	Promoción de la autonomía	0,993	0,990	0,044
		DOC4PMAT_APR	Creencias sobre la enseñanza de la matemática	0,984	0,978	0,042
		DOC4PMAT_RETRD	Retroalimentación del director	0,999	0,998	0,031

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR
4.º grado de primaria	Familia	FAM4PGEN_BIENEST	Malestar socioafectivo del estudiante	0,997	0,995	0,015
		FAM4PGEN_ESTEMO	Esteretotipos de género en las emociones	0,998	0,994	0,028
		FAM4PGEN_INVESC	Involucramiento parental educativo en la escuela	0,992	0,989	0,043
		FAM4PGEN_INVGEN	Involucramiento parental general	0,978	0,970	0,052
		FAM4PGEN_INVHOG	Involucramiento parental educativo en el hogar	0,996	0,994	0,036
		FAM4PGEN_SERVIE	Percepción sobre los servicios de la escuela	0,998	0,998	0,025
		FAM4PHSE_CRECAS	Creencias sobre el castigo físico en la crianza	0,998	0,997	0,029
		FAM4PLEC_PGUSTO	Gusto por la lectura (percepción de la familia)	1,000	1,000	0,001
		FAM4PMAT_PGUSTO	Gusto por la matemática (percepción de la familia)	1,000	0,999	0,012

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR
6.º grado de primaria	Docente de Comunicación	DOC6PLEC_CONFDD	Confianza director-docente	0,999	0,999	0,017
		DOC6PLEC_DESSOC	Deseabilidad social	0,982	0,977	0,053
		DOC6PLEC_EDUIN	Actitudes hacia la educación inclusiva	1,000	1,000	0,005
		DOC6PLEC_EVFORM	Creencias sobre la evaluación formativa	0,985	0,981	0,049
		DOC6PLEC_IEAPRE	Escuela orientada al desarrollo de aprendizajes	0,998	0,996	0,027
		DOC6PLEC_PRACLECT	Prácticas docentes en la enseñanza de la lectura	0,991	0,989	0,060
		DOC6PLEC_PROBIE	Percepción de situaciones problemáticas en la escuela relacionadas al estudiante	0,991	0,986	0,048
		DOC6PLEC_PROMAU	Promoción de la autonomía	0,994	0,991	0,039

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR
6.° grado de primaria	Docente de Matemática	DOC6PMAT_APR	Creencias sobre la enseñanza de la matemática	0,982	0,976	0,054
		DOC6PMAT_PREPMAT	Preparación del docente para la enseñanza de Matemática	0,995	0,993	0,045
		DOC6PMAT_RETRD	Retroalimentación del director	0,999	0,998	0,035
	Docente tutor	DOC6PHSE_AFRON	Estilos de afrontamiento	0 0,981	0 0,977	0 0,065
		DOC6PHSE_CONEMO	Autopercepción de la regulación de sus emociones	0,995	0,994	0,036
		DOC6PHSE_DESOCIAL	Deseabilidad social	0,978	0,973	0,057
		DOC6PHSE_SOEMO	Soporte emocional al estudiante	0,996	0,995	0,037
		FAM6PGEN_ACOSVIC	Acoso escolar al estudiante	0,996	0,994	0,043
	Familia	FAM6PGEN_BIENEST	Malestar socioafectivo del estudiante	0,992	0,984	0,024

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR
6.º grado de primaria	Familia	FAM6PGEN_ESTDES	Esteriotipos de género en el desempeño	0,997	0,993	0,035
		FAM6PGEN_ESTEMO	Esteriotipos de género en las emociones	0,999	0,996	0,021
		FAM6PGEN_INVESC	Involucramiento parental educativo en la escuela	0 0,991	0 0,988	0 0,044
		FAM6PGEN_INVGEN	Involucramiento parental general	0,972	0,963	0,065
		FAM6PGEN_INVHOG	Involucramiento parental educativo en el hogar	0,996	0,994	0,033
		FAM6PGEN_SERVIE	Percepción sobre los servicios de la escuela	0,995	0,994	0,036
		FAM6PHSE_TOLCAS	Tolerancia al castigo físico	0,998	0,995	0,032
		FAM6PLEC_PGUSTO	Gusto por la lectura (percepción de la familia)	1,000	1,000	0,008
		FAM6PMAT_PGUSTO	Gusto por la matemática (percepción de la familia)	1,000	0,999	0,014

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR
6.° grado de primaria	Estudiante	EST6PCCSS_PENSOC	Pensamiento social del estudiante	0,971	0,965	0,040
		EST6PCCSS_PENSODOC	Promoción del pensamiento social en clases	0,973	0,956	0,046
		EST6PGEN_DIFTEXT	Dificultad percibida de los textos	1,000	1,000	0,000
		EST6PGEN_ESTEST	Estrategias de estudio	0,954	0,945	0,045
		EST6PGEN_INTEXT	Interés por los textos en la escuela	0,996	0,993	0,022
		EST6PGEN_MOTL	Motivación hacia la lectura: tiempo libre	0,973	0,965	0,064
		EST6PGEN_OPARDOC	Disposición del docente para participación en clase	0,993	0,988	0,031
		EST6PGEN_OPAREST	Disposición del estudiante para participar en clase	0,987	0,961	0,030,
		EST6PGEN_SEGESC	Seguridad escolar	1,000	1,000,	0,000,

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR
6.º grado de primaria	Estudiante	EST6PGEN_SENPER	Sentido de pertenencia	0,996	0,994	0,030
		EST6PGEN_VIOLDOC	Manifestaciones de violencia por parte de los docentes	0,998	0,996	0,026
		EST6PGEN_VIOLEST	Manifestaciones de violencia por parte de los estudiantes	0,999	0,998	0,021
		EST6PHSE_CSEREG	Clima social escolar	0,999	0,998	0,013
		EST6PHSE_EAA	Estrés académico adolescente	0,990	0,985	0,045
		EST6PLEC_ACTCOG	Activación cognitiva en clases de Comunicación	0,992	0,989	0,023
		EST6PMAT_ACTCOG	Activación cognitiva - Matemática	0,980	0,972	0,038
		EST6PMAT_ANSIE	Ansiedad ante la matemática	0,995	0,993	0,037
		EST6PMAT_DISRP	Disrupción en clase - Matemática	0,998	0,996	0,018

Grado	Actor	Código de constructo	Nombre	CFI	TLI	SRMR
6.º grado de primaria	Estudiante	EST6PMAT_MOTMAT	Motivación hacia la Matemática	0,997	0,995	0,027
		EST6PHSE_APACAD	Clima de apoyo académico	0,999	0,998	0,019
		EST6PHSE_BIESUB	M Bienestar subjetivo	0,998	0,995	0,017
		EST6PHSE_DESEA	Deseabilidad Social	0,962	0,949	0,056
		EST6PHSE_ESTPA	Estilos parentales	0,984	0,981	0,053
		EST6PHSE_LEY	Transgresión a la ley	0 1,000	0 1,000	0 0,005
		EST6PHSE_PROSO	Conducta prosocial	0,989	0,986	0,052
		EST6PHSE_SOEMO	Soporte socioemocional del docente	0,997	0,996	0,027

B. Para constructos formativos

El análisis de constructos formativos fue realizado por medio de Análisis de Componentes Principales (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:

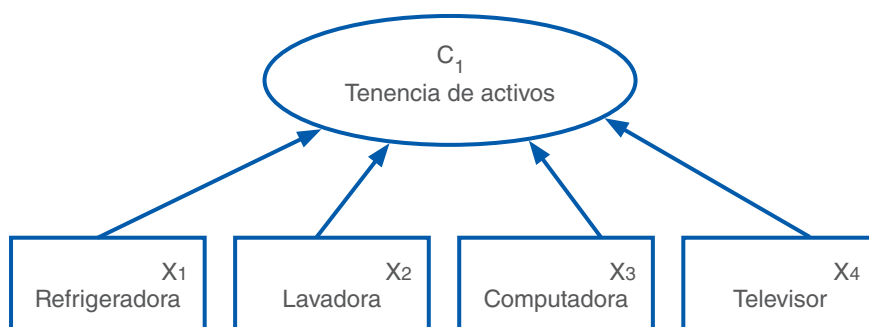
$$c_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m$$

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 2024, 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 fueron calculados a través de matrices de correlaciones policóricas con los ítems para cada indicador. Las correlaciones al interior de cada dimensión en la ENLA 2024 fueron positivas. Todas estas estimaciones se realizaron utilizando el paquete psych (Revelle, 2024) del lenguaje R (R Core Team) 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
4.° y 6.° grado de primaria	Director	DIRPGEN_DECIAC	Inclusión de comunidad educativa en decisiones escolares	52,06
		DIRP_DISPMA	Disponibilidad de materiales educativos textuales	59,79
		DIRP_DISPMAT	Disponibilidad de materiales educativos	63,78
	Familia	ISE	Índice socioeconómico ³	59,92
4.° grado de primaria	Familia	FAM4PGEN_DESINF	Desarrollo infantil temprano	49,22
6° grado de primaria	Docente tutor	DOC6PHSE_ACOMPA	Facilidad para el acompañamiento socioemocional y académico	56,66
		DOC6PHSE_DESAHSE	Promoción del desarrollo de las HSE en la escuela	71,33
		DOC6PHSE_PERCVIO	Percepción de violencia entre estudiantes	62,51
		DOC6PHSE_RIESGO	Percepción de conductas de riesgo en los estudiantes	62,99
		DOC6PHSE_TOE	Facilidad para abordar temas en TOE	58,99

³ Tal como se señaló el ISE está compuesto de por otras cuatro medidas que fueron derivadas de un análisis de componentes principales cuyas varianzas explicadas fueron las siguientes: 1) materiales de la vivienda (Var. Exp. = 72,58), 2) servicios básicos en el hogar (Var. Exp. = 79,59), 3) activos en el hogar (Var. Exp. = 60,49) y 4) acceso a otros servicios (Var. Exp. = 68,57).

Grado	Actor	Código de constructo	Nombre	Varianza explicada
6° grado de primaria	Estudiante	EST6PHSE_RIESGO	Conductas de riesgo	65,93
		EST6PHSE_RIFDIF	Relaciones intrafamiliares: dificultades	57,39
		EST6PHSE_RIFEX	Relaciones intrafamiliares: expresión	74,73
		EST6PHSE_RIFUN	Relaciones intrafamiliares: unión y apoyo	70,92
		EST6PHSE_VIOLE	Violencia escolar	79,95

Para el caso particular del índice socioeconómico, este se construyó a partir de los siguientes indicadores:

- 1) materiales de la vivienda,
- 2) servicios básicos en el hogar,
- 3) activos en el hogar,
- 4) acceso a otros servicios y
- 5) el máximo nivel educativo de los padres.

Los cuatro primeros fueron derivados de un análisis de componentes principales, mientras que el último se construyó a partir de los años de estudio promedio que implicaría cada uno de los niveles educativos. Un mayor detalle de este procedimiento se puede encontrar en Minedu (2018). Para el cálculo de las variables que conforman el índice socioeconómico se utilizaron en conjunto las bases de datos de las familias de 4.° grado de primaria y 6.° grado de primaria.

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:

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

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 compRelSEM 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
4.º y 6.º grado de primaria	Director	DIRP_COLABDOC	Colaboración entre docentes		0,842
		DIRPGEN_SITPROBR	Situaciones problemáticas en la escuela	Falta de recursos	0,976
		DIRPGEN_SITPROBD	Situaciones problemáticas en la escuela	Docentes	0,890
		DIRP_AUTIE	Autonomía para realizar mejoras en la escuela		0,896
		DIRP_SATUGELDOC	Satisfacción con el apoyo recibido por la UGEL	Apoyo a docentes	0,875
		DIRP_SATUGELDIR	Satisfacción con el apoyo recibido por la UGEL	Apoyo a directores	0,948
		D DIRP_LIDUGCONF	Liderazgo educativo de la UGEL	Fomento de la confianza	0,913
		DIRP_LIDUGCOL	Liderazgo educativo de la UGEL	Fomento de la colaboración	0,882

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
4.º y 6.º grado de primaria	Director	DIRP_ENSHSE	Creencias sobre la enseñanza de las HSE		0,688
		DIRP_CREAUTO	Creencias sobre la autorregulación		0,660
		DIRP_UTILLEC	Percepción de utilidad de materiales educativos de Comunicación		0,582
		DIRP_UTILMAT	Percepción de utilidad de materiales educativos de Matemática		0,603
4.º grado de primaria	Docente de Comunicación	DOC4PLEC_CONFDD	Confianza director-docente		0,878
		DOC4PLEC_CRERLEC	Creencias sobre la recursividad de la escritura		0,718
		DOC4PLEC_EVFORMCR	Creencias sobre la evaluación formativa	Criterios erróneos de evaluación	0,785
		DOC4PLEC_EVFORMRE	Creencias sobre la evaluación formativa	Creencias erróneas sobre retroalimentación	0,774
		DOC4PLEC_IEAPRE	Escuela orientada al desarrollo de aprendizajes		0,851
		DOC4PLEC_PRACLECDEC	Prácticas docentes en la enseñanza de la lectura	Decodificación	0,692
		DOC4PLEC_PRACLECREP	Prácticas docentes en la enseñanza de la lectura	Repetición	0,942

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
4.º grado de primaria	Docente de Comunicación	DOC4PLEC_PRACLECTSIG	Prácticas docentes en la enseñanza de la lectura	Significado	0,796
		DOC4PLEC_PREPCOM	Preparación del docente para la enseñanza de Lectura		0,950
		DOC4PLEC_PROBIE	Percepción de situaciones problemáticas en la escuela		0,864
		DOC4PLEC_PROMAU	Promoción de la autonomía		0,855
	Docente de Matemática	DOC4PMAT_APRACT	Creencias sobre la enseñanza de la matemática	Enfoque de aprendizaje activo	0,654
		DOC4PMAT_APRTRA	Creencias sobre la enseñanza de la matemática	Enfoque tradicional	0,734
		DOC4PMAT_RETRDG	Retroalimentación del director	Reuniones grupales	0,837
		DOC4PMAT_RETRDP	Retroalimentación del director	Retroalimentación individual	0,927
	Familia	FAM4PGEN_BIENEST	Malestar socioafectivo del estudiante		0,663
		FAM4PGEN_ESTEMO	Esteretotipos de género en las emociones		0,809
		FAM4PGEN_INVESC_ACT	Involucramiento parental educativo en la escuela	Participación en las actividades de la escuela	0,860

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
4.° grado de primaria	Familia	FAM4PGEN_INVESC_DÖC	Involucramiento parental educativo en la escuela	Interacción con los docentes	0,824
		FAM4PGEN_INVGENAC	Involucramiento parental general	Realizar actividades	0,702
		FAM4PGEN_INVGENCO	Involucramiento parental general	Conversar	0,712
		FAM4PGEN_INVHOG	Involucramiento parental educativo en el hogar		0,890
		FAM4PGEN_SERVIE	Percepción sobre los servicios de la escuela		0,887
		FAM4PHSE_CRECAS	Creencias sobre el castigo físico en la crianza		0,886
		FAM4PLEC_PGUSTO	Gusto por la lectura (percepción de la familia)		0,796
		FAM4PMAT_PGUSTO	Gusto por la matemática (percepción de la familia)		0,785
6.° grado de primaria	Docente de Comunicación	DOC6PLEC_CONFDD	Confianza director-docente		0,874
		DOC6PLEC_DESSOC	Deseabilidad social		0,661
		DOC6PLEC_EDUIN	Actitudes hacia la educación inclusiva		0,852

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
6.º grado de primaria	Docente de Comunicación	DOC6PLEC_EVFORMCR	Creencias sobre la evaluación formativa	Criterios erróneos de evaluación	0,802
		DOC6PLEC_EVFORMRE	Creencias sobre la evaluación formativa	Creencias erróneas sobre retroalimentación	0,791
		DOC6PLEC_IEAPRE	Escuela orientada al desarrollo de aprendizajes		0,853
		DOC6PLEC_PRACLECDEC	Prácticas docentes en la enseñanza de la lectura	Decodificación	0,661
		DOC6PLEC_PRACLECREP	Prácticas docentes en la enseñanza de la lectura	Repetición	0,896
		DOC6PLEC_PRACLECTSIG	Prácticas docentes en la enseñanza de la lectura	Significado	0,790
		DOC6PLEC_PROBIE	Percepción de situaciones problemáticas en la escuela relacionadas al estudiante		0,857
		DOC6PLEC_PROMAU	Promoción de la autonomía		0,856
	Docente de Matemática	DOC6PMAT_APRACT	Creencias sobre la enseñanza de la matemática	Enfoque de aprendizaje activo	0,753
		DOC6PMAT_APRTRA	Creencias sobre la enseñanza de la matemática	Enfoque tradicional	0,772
		DOC6PMAT_PREPMAT	Preparación del docente para la enseñanza de Matemática		0,895

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
6.º grado de primaria	Docente de Matemática	DOC6PMAT_RETIN	Retroalimentación del director	Retroalimentación individual	0,940
		DOC6PMAT_RETRG	Retroalimentación del director	Reuniones grupales	0,870
	Docente tutor	DOC6PHSE_AFRONEE	Estilos de afrontamiento	Evitacion emocional	0,847
		DOC6PHSE_AFRONSP	Estilos de afrontamiento	Solución de problemas	0,854
		DOC6PHSE_CONEMOC	Auto percepción de la regulación de sus emociones	Claridad emocional (compresión)	0,840
		DOC6PHSE_CONEMOP	Auto percepción de la regulación de sus emociones	Atencion a las emociones (percepción)	0,853
		DOC6PHSE_CONEMOR	Auto percepción de la regulación de sus emociones	Reparación emocional (regulación)	0,808
		DOC6PHSE_DESOCIAL	Deseabilidad social		0,681
		DOC6PHSE_SOEMOCP	Soporte emocional al estudiante	Consideración por las perspectivas de los adolescentes	0,808
		DOC6PHSE_SOEMOSD	Soporte emocional al estudiante	Sensibilidad docente	0,834

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
6.º grado de primaria	Estudiante	EST6PCCSS_PENSOCNE	Pensamiento social del estudiante	Negativo	0,261
		EST6PCCSS_PENSOCPO	Pensamiento social del estudiante	Positivo	0,780
		EST6PCCSS_PENSODON	Promoción del pensamiento social en clases	Negativo	0,624
		EST6PCCSS_PENSODOP	Promoción del pensamiento social en clases	Positivo	0,601
		EST6PGEN_DIFTEXT	Dificultad percibida de los textos		0,560
		EST6PGEN_ESTESTRE	Estrategias de estudio	Estrategias de estudio de baja elaboración	0,495
		EST6PGEN_ESTESTSI	Estrategias de estudio	Estrategias de estudio de alta elaboración	0,737
		EST6PGEN_INTEXT	Interés por los textos en la escuela		0,735
		EST6PGEN_MOTLAU	Motivación hacia la lectura: tiempo libre	Motivación autónoma	0,797
		EST6PGEN_MOTLCO	Motivación hacia la lectura: tiempo libre	Motivación controlada	0,775
		EST6PGEN_OPARDOCN	Diposición del docente para participación en clase	Negativo	0,679

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
6.º grado de primaria	Estudiante	EST6PGEN_OPARDOCP	Diposición del docente para participación en clase	Positivo	0,722
		EST6PGEN_OPPAREST	Disposición del estudiante para participar en clase		0,590
		EST6PGEN_SEGESC	Seguridad escolar		0,609
		EST6PGEN_SENPER	Sentido de pertenencia		0,834
		EST6PGEN_VIOLDOC	Manifestaciones de violencia por parte de los docentes		0,857
		EST6PGEN_VIOLEST	Manifestaciones de violencia por parte de los estudiantes		0,875
		EST6PHSE_CSEREG	Clima social escolar		0,746
		EST6PHSE_EAACT	Estrés académico adolescente	Carga de trabajo	0,800
		EST6PHSE_EAAPN	Estrés académico adolescente	Preocupación por las notas	0,711
		EST6PLEC_ACTCOG	Activación cognitiva en clases de Comunicación		0,736
		EST6PMAT_ACTCOG	Activación cognitiva - Matemática		0,735

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
6.º grado de primaria	Estudiante	EST6PMAT_ ANSIE	Ansiedad ante la matemática		0,903
		EST6PMAT_ DISRP	Disrupción en clase - Matemática		0,810
		EST6PMAT_ MOTMAT	Motivación hacia la Matemática		0,832
		EST6PHSE_ APACAD	Clima de apoyo académico		0,880
		EST6PHSE_ BIESUB	Bienestar subjetivo		0,742
		EST6PHSE_ DESEA	Deseabilidad Social		0,646
		EST6PHSE_ ESTPA_COM	Estilos parentales	Compromiso	0,862
		EST6PHSE_ ESTPA_CP	Estilos parentales	Control psicológico	0,711
		EST6PHSE_ LEY	Transgresión a la ley		0,827
6.º grado de primaria	Docente	EST6PHSE_ PROSO	Conducta prosocial		0,926
		EST6PHSE_ SOEMO	Soporte socioemocional del docente		0,901

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
6.º grado de primaria	Familia	FAM6PGEN_ACOSVIC	Acoso escolar al estudiante		0,886
		FAM6PGEN_BIENEST	Malestar socioafectivo del estudiante		0,631
		FAM6PGEN_ESTDES	Esteretipos de género en el desempeño		0,906
		FAM6PGEN_ESTEMO	Esteretipos de género en las emociones		0,836
		FAM6PGEN_INVESC_ACT	Involucramiento parental educativo en la escuela	Involucramiento parental: participación en las actividades de la escuela	0,857
		FAM6PGEN_INVESC_DOC	Involucramiento parental educativo en la escuela	Involucramiento parental: interacción con los docentes	0,830
		FAM6PGEN_INVGENAC	Involucramiento parental general	Realizar actividades	0,735
		FAM6PGEN_INVGENCO	Involucramiento parental general	Conversar	0,772
		FAM6PGEN_INVHOG	Involucramiento parental educativo en el hogar		0,887
		FAM6PGEN_SERVIE	Percepción sobre los servicios de la escuela		0,895
		FAM6PHSE_TOLCAS	Tolerancia al castigo físico		0,876

Grado	Actor	Código de constructo	Nombre general	Sub escala	Omega
6.º grado de primaria	Familia	FAM6PLEC_PGUSTO	Gusto por la lectura (percepción de la familia)		0,748
		FAM6PMAT_PGUSTO	Gusto por la matemática (percepción de la familia)		0,780

5.4 Elaboración de puntajes

Una vez estimada 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 les 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 Model), 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.

Diseño, construcción y procesamiento de datos de Habilidades Socioemocionales

Capítulo 6

Capítulo 6

Diseño, construcción y procesamiento de datos de Habilidades Socioemocionales

En la Evaluación Nacional de Logros del Aprendizaje (ENLA) 2024, promovida por el Ministerio de Educación del Perú, se ha incluido la medición de habilidades socioemocionales como parte de un enfoque integral para comprender y fortalecer el bienestar de los estudiantes en el país.

Las habilidades socioemocionales son entendidas como aquellas aptitudes o destrezas de una persona para identificar, expresar y gestionar sus emociones, comprender las de los demás, demostrar empatía, tomar decisiones responsables, y relacionarse saludable y satisfactoriamente con los otros. Estas habilidades influyen en el desempeño de las personas y son transferibles a distintos contextos, como el familiar, el escolar, el laboral y el ciudadano¹. Las habilidades socioemocionales se pueden desarrollar a través de experiencias de aprendizaje formales (actividades intencionadas) e informales (actividades cotidianas). Ambos tipos de experiencias van a permitir generar estrategias para afrontar desafíos presentes en la vida cotidiana y en las interacciones con los demás.

En ese sentido, desde el Currículo Nacional de Educación Básica (CNEB; Ministerio de Educación del Perú [Minedu], 2016), documento marco que guía la política educativa nacional, reconoce que la acción educativa contempla el acompañamiento formativo de capacidades individuales, cognitivas y socioemocionales. Adicionalmente, el Proyecto Educativo Nacional (Consejo Nacional de Educación, 2020) contempla al bienestar socioemocional como un propósito fundamental para la ciudadanía plena, pues implica manejar las emociones de manera adaptativa y fomentar la convivencia saludable, contribuyendo al bienestar colectivo.

Asu vez, el Minedu publicó el marco de Desarrollo de las Habilidades Socioemocionales en el marco de la Tutoría y Orientación Educativa (Minedu, 2021), en el que se contemplan trece HSE relevantes que están alineadas con las competencias explicitadas en el CNEB (Minedu, 2016). Actualmente, se están priorizando diez HSE, las cuales son: Autoconcepto, Conciencia emocional, Autocuidado, Empatía, Comunicación asertiva, Regulación emocional, Toma de decisiones responsable, Resolución de conflictos, Trabajo en equipo y Conciencia social.

¹ Esta definición ha sido extraída del documento *Desarrollo de las habilidades socioemocionales en el marco de la Tutoría y Orientación Educativa* (Ministerio de Educación, 2021).

Considerando el escenario de la post pandemia, en el 2022, se promulga el Decreto Supremo N.º 013-2022-Minedu el cual aprueba los Lineamientos para la promoción del bienestar socioemocional de las y los estudiantes de la Educación Básica (Minedu, 2022). De la misma forma, de acuerdo con el Resolución Ministerial N.º 556-2024-Minedu se identificó la necesidad de una atención especial a la salud mental, a la gestión de las emociones de los estudiantes (Minedu, 2024). En ese sentido, en la “Norma Técnica para el Año Escolar en las Instituciones y programas educativos públicos y privados de la Educación Básica para el Año 2025”, una de sus prioridades es promover el bienestar socioemocional en las actividades desarrolladas a lo largo del año escolar, así como desarrollar una educación inclusiva, intercultural y equitativa.

6.1 Conceptos clave

Dentro de las habilidades socioemocionales, la autorregulación ocupa un lugar fundamental, dado su impacto transversal en el aprendizaje académico, la convivencia escolar y el bienestar personal.

En el marco de la ENLA 2024, la autorregulación, ha sido definida como la habilidad que comprende la gestión (regulación) de las emociones propias ante diversas situaciones retadoras de la vida diaria. Esta habilidad nos permite reconocer la relación entre emoción, cognición (pensamiento) y comportamiento (conducta), y, a partir de ello, generar estrategias para alcanzar determinadas metas².

La habilidad de autorregulación, tal como fue evaluada en la ENLA 2024, considera tres aspectos fundamentales. En primer lugar, se encuentra la conciencia emocional, que hace referencia a la capacidad de reconocer, identificar y nombrar las propias emociones. Este aspecto es clave porque permite que los estudiantes puedan poner en palabras lo que sienten, generando así un primer nivel de conocimiento sobre sus estados emocionales. En segundo lugar, se consideró la regulación emocional, entendida como la capacidad de gestionar la intensidad y duración de las emociones para promover el bienestar y afrontar situaciones adversas. Esta dimensión es especialmente relevante en el contexto escolar, donde los estudiantes enfrentan presiones asociadas al rendimiento académico, la interacción social y el respeto de normas. La regulación emocional permite que los estudiantes respondan de forma adaptativa a estas situaciones, reduciendo niveles de estrés, frustración o ansiedad, y favoreciendo una actitud resiliente (Gross, 2015).

Finalmente, el tercer aspecto es la regulación conductual, la cual se refiere a la capacidad de desarrollar y mantener, de manera flexible, un comportamiento planificado con el fin de alcanzar nuestros objetivos en diversas situaciones.

² Esta definición ha sido elaborada atendiendo a los lineamientos del CNEB. Asimismo, se tuvo en consideración lo referido por organismos como el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) de la Unesco y la Agencia de la Calidad de la Educación de Chile.

6.2 Procedimiento para la elaboración del instrumento

El instrumento utilizado en la ENLA 2024 para evaluar esta habilidad fue construido tomando como base dos fuentes principales, la versión abreviada del Self-Regulation Questionnaire (SRQ) desarrollada por Pichardo et al. (2014) y la escala de Regulación Emocional elaborada por la Mesa Regional de Cooperación Técnica sobre Competencias Técnicas y Socioemocionales (MESACTS, 2020). Es pertinente señalar que la escala SRQ ha sido previamente utilizada por la Oficina de Medición de la Calidad de los Aprendizajes durante la pandemia en el Estudio Virtual de Aprendizajes de 2021 y se retomó de forma presencial en la Evaluación Muestral de Estudiantes de 2022, con el propósito de evaluar la autorregulación conductual en estudiantes.

Por tanto, se puede señalar que el instrumento utilizado en la ENLA 2024 tuvo como base dos escalas previamente desarrolladas, ambas escalas aportaron componentes clave para abordar la autorregulación, con el objetivo de que el instrumento final refleje los tres aspectos que componen la autorregulación (conciencia emocional, regulación emocional y regulación conductual).

La experiencia acumulada en el uso y obtención de evidencias de validez de estas herramientas respalda su pertinencia y utilidad en el ámbito educativo. Por ejemplo, sobre la escala SRQ, estudios como el de Pichardo et al. (2018) respaldan la validez y utilidad del SRQ en contextos educativos, al identificar factores relevantes como la planificación, la perseverancia y la toma de decisiones.

Asimismo, las evidencias de validez incluyeron un proceso que contempló la revisión de expertos, una aplicación de un estudio piloto en diferentes regiones del país y el análisis psicométrico, lo cual contribuye a que los resultados obtenidos reflejen de manera adecuada la habilidad evaluada.

Por otro lado, el diseño del instrumento no solo consideró aspectos técnicos y psicométricos, sino también criterios pedagógicos y de desarrollo evolutivo, lo cual buscó favorecer su aplicabilidad en estudiantes del nivel primario. Los ítems fueron redactados con un lenguaje accesible y directo, y se buscó que estos representen experiencias escolares o propias de su vida cotidiana, para que los estudiantes puedan reflexionar con mayor claridad sobre su propio comportamiento emocional y conductual.

6.3. Análisis psicométricos para habilidades socioemocionales

Inicialmente, en el operativo de campo, se aplicaron los 36 ítems originales provenientes de las escalas desarrolladas por Pichardo et al. (2014) y la MESACTS (2020). Estos fueron respondidos mediante una escala de valoración tipo Likert de cinco puntos, que va desde *Totalmente en desacuerdo* hasta *Totalmente de acuerdo*. En este formato de respuesta, una mayor puntuación indica una mayor percepción

de autorregulación por parte del estudiante. Las respuestas de los estudiantes fueron sometidos a un análisis mediante modelos Rasch, que permitieron evaluar el ajuste de los ítems. Primero se aplicó el modelo de escala de valoración de Rasch (Rasch Rating Scale Model; Andrich, 1978) en su versión multidimensional, pero este no mostró un buen ajuste, habiendo correlaciones muy altas entre las tres supuestas dimensiones, lo cual hacía que no se las distinga con facilidad. Por ello se optó por una versión unidimensional.

Como resultado de este análisis unidimensional, se identificó que 14 de los 36 ítems no presentaban un ajuste adecuado, por lo que fueron descartados para los análisis posteriores. De esta manera, la escala final estuvo compuesta por 22 ítems.

6.3.1 Calibración final de los ítems

En la tabla 6.1, se presentan los resultados de la calibración de los ítems de la escala 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, que 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, que 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 la tabla 6.1, los errores de calibración son muy pequeños, lo que indica que la calibración de la dificultad fue bastante precisa. Por último, en la tabla se observa que el infit y el outfit de todos los ítems de la escala de HSE empleada en la ENLA 2024 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 6.1. Calibración de los ítems de la escala de HSE aplicada en 6.º grado de primaria

Ítem	Medida	Error	Infit	Outfit
HSECOND_01	-0,651	0,003	0,90	0,94
HSECOND_02	-0,660	0,003	0,91	0,94
HSECOND_04	-0,399	0,003	1,05	1,10
HSECOND_07	-1,515	0,004	1,10	1,02
HSECOND_08	-0,429	0,003	1,06	1,11
HSECOND_10	-0,232	0,003	0,99	1,08
HSECOND_11	-0,987	0,004	0,81	0,80
HSECOND_12	-0,450	0,003	0,99	1,03

Ítem	Medida	Error	Infit	Outfit
HSECOND_13	-0,358	0,003	1,04	1,10
HSECOND_14	-0,264	0,003	1,05	1,11
HSECONS_01	-0,729	0,004	0,84	0,82
HSECONS_03	-0,596	0,003	0,95	0,95
HSECONS_04	-0,700	0,004	0,92	0,91
HSECONS_05	-0,684	0,003	0,99	1,00
HSECONS_06	-0,448	0,003	1,23	1,27
HSECONS_08	-0,919	0,004	1,14	1,10
HSEREG_01	-0,725	0,004	0,90	0,89
HSEREG_02	-0,469	0,003	1,13	1,17
HSEREG_04	-0,649	0,003	1,21	1,22
HSEREG_05	-0,489	0,003	1,06	1,07
HSEREG_08	-0,460	0,003	1,09	1,11
HSEREG_09	-0,894	0,004	1,13	1,10

6.3.2 Estructura de umbrales

Una característica adicional del modelo de escala de valoración de Rasch (respecto del 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 que es común para todos los ítems. Los umbrales, también conocidos como umbrales de Andrich, se definen como los puntos en los que dos categorías de respuesta adyacentes tienen la misma probabilidad de ser elegidas. De esta manera, cuando el nivel de habilidad de un estudiante se ubica justo en el punto de un umbral, tiene la misma probabilidad de elegir una u otra de las dos categorías adyacentes. En ese sentido, si su nivel de habilidad supera ese umbral, es más probable que seleccione la categoría superior. Por ello, los umbrales marcan el punto en el que cambia la probabilidad de pasar de una categoría a otra. Es importante señalar que los umbrales deberían estar ordenados de menor a mayor, ya que se espera que las categorías más altas reflejen niveles más elevados en la medida, mientras que las más bajas correspondan a niveles más bajos³ (Linacre, 1999).

Por último, es relevante considerar que el orden de la estructura de los umbrales depende de las frecuencias observadas en cada categoría de respuesta. De manera que, por ejemplo, cuando una categoría intermedia tiene una mayor proporción de

³ 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).

respuestas que la siguiente categoría adyacente, podría obtenerse una formación de umbrales desordenados (Linacre, s.f.). A pesar de ello, una estructura desordenada de los umbrales, no implica un mal ajuste del modelo Rasch (Linacre, s.f.).

La Tabla 6.2 presenta la estructura de umbrales de la escala de autorregulación. Aunque los umbrales no se encuentran ordenados de manera creciente, dicho hallazgo no representa necesariamente un mal ajuste del modelo.

Tabla 6.2. Estructura de umbrales de la escala de HSE aplicada en 6.º grado de primaria

Escala	Primer umbral		Segundo umbral		Tercer umbral		Cuarto umbral	
	Medida	Error	Medida	Error	Medida	Error	Medida	Error
Autorregulación	-0,885	0,002	-0,277	0,002	-0,343	0,002	1,505	-

6.3.3 Evidencias de validez vinculadas a la estructura interna

En la tabla 6.3, se presentan los resultados del análisis de componentes principales de los residuos del modelo Rasch. A partir de ello, se concluye que los datos provenientes de aplicar la escala de autorregulación utilizada en la ENLA 2024 no presentan serias evidencias en contra del supuesto de unidimensionalidad, por lo cual es posible asumir que la varianza observada se explica básicamente por la presencia de un rasgo latente dominante.

Tabla 6.3. Análisis de unidimensionalidad de la medida de HSE derivada de la escala aplicada en 6.º grado de primaria

Escala	Medidas Rasch		ACPR	
	Autovalor	Varianza explicada (%)	Autovalor	Varianza explicada (%)
Autorregulación	9,96	31,2	2,08	6,5

6.3.4 Confiabilidad

La confiabilidad se calculó utilizando la estimación EAP (*expected a posteriori*) de las medidas de las personas. En relación con la escala de HSE aplicada en 6.º grado de primaria, esta mostró medidas con una confiabilidad de 0,89, la cual está por encima del punto de corte mínimo recomendable (0,70).

6.4 Presentación de resultados

Las medidas de habilidad de las personas fueron estimadas con la metodología de valores plausibles⁴. Se utilizaron cinco valores plausibles por cada persona evaluada, ya que se ha demostrado que esta metodología permite obtener estimaciones más precisas de la variabilidad de la habilidad de los estudiantes a nivel poblacional, en comparación con otras estimaciones puntuales como el *Maximum Likelihood Estimation* o el *Weighted Maximum Likelihood* (Wu, 2005). Para la estimación de tales valores plausibles se empleó una regresión latente, el cual es un modelo que predice dichos valores a partir de covariables observables, con el fin de aproximarse a la habilidad verdadera de los estudiantes. En este caso, se consideraron como predictores el sexo, el tipo de gestión de la institución educativa y el área geográfica.

Para la presentación de los resultados descriptivos se optó por expresarlos en porcentajes con el objetivo de facilitar su interpretación y su comunicación a la comunidad educativa. De esta manera, se aplicó un umbral de inclusión de casos basado en una presencia mínima de valores perdidos en las respuestas del cuestionario; es decir, solo se incluyeron en el análisis descriptivo los cuestionarios de los estudiantes que respondieron al menos una cantidad mínima de ítems. Específicamente, se excluyeron del análisis aquellos estudiantes que no respondieron al menos el 75 % de los ítems; es decir, se consideró un mínimo de 17 de los 22 ítems totales. Posteriormente, se recodificaron las respuestas de la escala tipo Likert de cinco puntos (*Totalmente en desacuerdo*, *En desacuerdo*, *Ni de acuerdo ni en desacuerdo*, *De acuerdo* y *Totalmente de acuerdo*), agrupando las dos categorías más altas (*De acuerdo* y *Totalmente de acuerdo*) en una única categoría denominada “respuesta positiva”. A partir de esta nueva codificación, se calculó, para cada ítem, el porcentaje de estudiantes que emitieron una respuesta positiva. Finalmente, se estimó la proporción promedio de respuestas positivas en el total de la escala.

⁴ Los valores plausibles son estimaciones posibles del puntaje real de una persona en una prueba, calculadas a partir de su desempeño y de un modelo estadístico. En vez de dar un solo puntaje, se generan varios valores probables para representar mejor la incertidumbre de la medición.

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>
- 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>
- Bollen, K. A., & Diamantopoulos, A. (2017). In defense of causal-formative indicators: A minority report. *Psychological Methods*, 22(3), 581–596. <https://doi.org/10.1037/met0000056>
- Bond, T. G., Yan, Z. & 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., & 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>
- 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>

- Embretson, S., & 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.
- Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*(1), 1–26. <https://doi.org/10.1080/1047840X.2014.940781>
- Hambleton, R., Swaminathan, H. & 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., & 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., & 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. & 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., & 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>
- Lewis, D. M., Mitzel, H. C. & 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>
- Linacre, J. M. (s.f.). *Andrich thresholds: Disordered rating or partial credit structures*. Recuperado el 23 de abril de 2025, de <https://winsteps.com/winman/disorder.htm>
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Educational Testing Service. <https://www.ets.org/Media/Research/pdf/LIVINGSTON2ed.pdf>
- 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>
- Mesa Regional de Cooperación Técnica sobre Competencias Transversales y Socioemocionales [MESACTS]. (2020, febrero). *Escala de regulación emocional: Informe técnico*. MESACTS.
- 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>
- Ministerio de Educación. (2016). *Currículo Nacional de la Educación Básica*. <https://minedu.gob.pe/curriculo/pdf/curriculo-nacional-de-la-educacion-basica.pdf>
- Ministerio de Educación. (2018). *Desafíos en la medición y el análisis del estatus socioeconómico de los estudiantes peruanos*. Oficina de Medición de la Calidad de los Aprendizajes. <http://umc.minedu.gob.pe/desafios-en-la-medicion-y-el-analisis-del-estatus-socioeconomico-de-los-estudiantes-peruanos/>
- Ministerio de Educación. (2021). *Desarrollo de las habilidades socioemocionales en el marco de la Tutoría y Orientación Educativa*. <https://w3.perueduca.pe/documents/10179/e9edd1ac-e713-43dc-bf53-fa5ac6653650>

- Ministerio de Educación. (2022). *Decreto Supremo N.º 013-2022-Minedu, que aprueba los Lineamientos para la promoción del bienestar socioemocional de las y los estudiantes de la Educación Básica*. <https://www.grade.org.pe/crear/archivos/DS-013-2022-minedu.pdf>
- Ministerio de Educación. (2024). *Resolución Ministerial N.º 556-2024-Minedu: Norma Técnica para el Año Escolar en las instituciones y programas educativos públicos y privados de la Educación Básica para el año 2025*. <https://www.gob.pe/institucion/minedu/normas-legales/6150505-556-2024-minedu>
- Moreno, R., Martínez, R. J. & 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., & Bernstein, I. (1995). *Teoría psicométrica* (3rd ed.). McGraw-Hill.
- 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.
- Pichardo, C., Justicia, F., de la Fuente, J., Martínez-Vicente, J. M. & Berbén, A. B. G. (2014). Factor structure of the Self-Regulation Questionnaire (SRQ) at Spanish universities. *The Spanish Journal of Psychology*, 17, Artículo E62.
- Pichardo, M. C., Cano, F., Garzón-Umerenkova, A., de la Fuente, J., Peralta-Sánchez, F. J. & Amate-Romera, J. (2018). Self-Regulation Questionnaire (SRQ) in Spanish adolescents: Factor structure and Rasch analysis. *Frontiers in Psychology*, 9, Artículo 1370. <https://doi.org/10.3389/fpsyg.2018.01370>
- 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., & 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., & 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., & 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., & 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., & Stone, M. (1998). *Diseño de mejores pruebas*. CENEVAL.
- Wright, B. D., & 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

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

Capacidad	Total	Porcentaje de la prueba
Obtiene información del texto escrito.	32	32,3 %
Infiere e interpreta información del texto escrito.	45	45,5 %
Reflexiona y evalúa la forma, el contenido y el contexto del texto escrito.	22	22,2 %
TOTAL	99	100 %

Tabla A.2 Distribución de textos en la prueba

Tipo textual	Género textual	Cantidad por tipo de texto	Porcentaje por tipo de texto
Narrativo	Cuento	4	23,5 %
	Historieta		
	Noticia		
	Anécdota		
Descriptivo	Artículo enciclopédico	4	23,5 %
	Artículo enciclopédico		
	Artículo periodístico		
	Infografía		
Instructivo	Manual de procedimientos	3	17,6 %
	Manual de procedimientos		
	Manual de procedimientos		
Argumentativo	Infografía	3	17,6 %
	Afiche		
	Artículo de opinión		
Expositivo	Artículo de divulgación	3	17,6 %
	Artículo de divulgación		
	Artículo de divulgación		
Total		17	100%

Tabla A.3 Distribución de ítems en la prueba

Tipo textual	Género textual	Cantidad de ítems por cada texto	Cantidad de ítems por tipo de texto	Porcentaje de ítems por tipo de texto
Narrativo	Cuento	6	21	21,2 %
	Historieta	6		
	Noticia	6		
	Anécdota	3		
Descriptivo	Artículo enciclopédico	6	24	24,2 %
	Artículo enciclopédico	6		
	Artículo periodístico	6		
	Infografía	6		
Instructivo	Manual de procedimientos	6	18	18,1 %
	Manual de procedimientos	6		
	Manual de procedimientos	6		
Argumentativo	Infografía	6	20	20,2 %
	Afiche	6		
	Artículo de opinión	8		
Expositivo	Artículo de divulgación	8	16	16,2 %
	Artículo de divulgación	2		
	Artículo de divulgación	6		
Total			99	100 %

Tabla B.1 Matemática ENLA 2024. Distribución de ítems por competencias en la prueba

Competencias	Cantidad de ítems	Porcentaje
Resuelve problemas de cantidad	47	42,7 %
Resuelve problemas de regularidad, equivalencia y cambio	23	20,9 %
Resuelve problemas de forma, movimiento y localización	23	20,9 %
Resuelve problemas de gestión de datos e incertidumbre	17	15,5 %
Total	110	100 %

Tabla B.2 Matemática ENLA 2024. Distribución de ítems por capacidades en la prueba

Competencia	Capacidades	Cantidad de ítems	Porcentaje
Resuelve problemas de cantidad	Traduce cantidades a expresiones numéricas.	27	24,5 %
	Comunica su comprensión sobre los números y las operaciones.	13	11,8 %
	Usa estrategias y procedimientos de estimación y cálculo.	4	3,6 %
	Argumenta afirmaciones sobre las relaciones numéricas y las operaciones.	3	2,7 %
Resuelve problemas de regularidad, equivalencia y cambio	Traduce datos y condiciones a expresiones algebraicas y gráficas.	2	1,8 %
	Comunica su comprensión sobre las relaciones algebraicas.	4	3,6 %
	Usa estrategias y procedimientos para encontrar equivalencias y reglas generales.	11	10,0 %
	Argumenta afirmaciones sobre las relaciones de cambio y equivalencia.	6	5,5 %
Resuelve problemas de forma, movimiento y localización	Modela objetos con formas geométricas y sus transformaciones.	3	2,7 %
	Comunica su comprensión sobre las formas y relaciones geométricas.	13	11,8 %
	Usa estrategias y procedimientos para orientarse en el espacio.	5	4,5 %
	Argumenta afirmaciones sobre relaciones geométricas.	2	1,8 %

Competencia	Capacidades	Cantidad de ítems	Porcentaje
Resuelve problemas de gestión de datos e incertidumbre	Representa datos con gráficos y medidas estadísticas o probabilísticas.	1	0,9 %
	Comunica su comprensión de los conceptos estadísticos y probabilísticos.	9	8,2 %
	Usa estrategias y procedimientos para recopilar y procesar datos.	4	3,6 %
	Sustenta conclusiones o decisiones con base en la información obtenida.	3	2,7 %
Total		102	110

El total ha sido redondeado a un decimal. Por ello, la suma total en esta tabla no es exactamente 100 %.

Tabla B.3 Matemática ENLA 2024. Distribución de ítems por contextos en la prueba

Contextos	Cantidad de ítems	Porcentaje
Intramatemático	12	10,9 %
Extramatemático	98	89,1 %
Total	110	100 %

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

Capacidad	Total	Porcentaje de la prueba
Obtiene información del texto escrito	27	25,2 %
Infiere e interpreta información del texto	48	44,9 %
Reflexiona y evalúa la forma, el contenido y el contexto del texto	32	29,9 %
TOTAL	107	100 %

Tabla C.2 Distribución de textos en la prueba

Tipo de texto	Género	Cantidad por tipo de texto	Porcentaje por tipo de texto
Narrativo	Cuento	3	16,7 %
	Cuento		
	Noticia		
Instructivo	Manual de procedimientos	3	16,7 %
	Manual de procedimientos		
	Manual de procedimientos		
Descriptivo	Artículo enciclopédico	4	22,2 %
	Artículo enciclopédico		
	Artículo enciclopédico		
	Descripción de personaje		
Argumentativo	Artículo de opinión	4	22,2 %
	Artículo de opinión		
	Carta		
	Infografía		
Expositivo	Artículo de divulgación	4	22,2 %
	Artículo de divulgación		
	Artículo de divulgación		
	Artículo de divulgación		
TOTAL		18	100 %

Tabla C.3 Distribución de ítems por texto en la prueba

Tipo de texto	Género	Cantidad de ítems por texto	Cantidad de ítems por tipo de texto	Porcentaje por tipo de texto
Narrativo	Cuento	6	18	16,8 %
	Cuento	6		
	Noticia	6		
Instructivo	Manual de procedimientos	6	18	16,8 %
	Manual de procedimientos	6		
	Manual de procedimientos	6		
Descriptivo	Artículo enciclopédico	6	23	21,5 %
	Artículo enciclopédico	6		
	Artículo enciclopédico	6		
	Retrato	5		
Argumentativo	Artículo de opinión	6	24	22,4 %
	Artículo de opinión	6		
	Carta	6		
	Infografía	6		
Expositivo	Artículo de divulgación	6	24	22,4 %
	Artículo de divulgación	6		
	Artículo de divulgación	6		
	Artículo de divulgación	6		
TOTAL		107	107	100 %

Tabla D.1 Matemática ENLA 2024. Distribución de ítems por competencias en la prueba

Competencias	Cantidad de ítems	Porcentaje
Resuelve problemas de cantidad	38	37,3 %
Resuelve problemas de regularidad, equivalencia y cambio	24	23,5 %
Resuelve problemas de forma, movimiento y localización	20	19,6 %
Resuelve problemas de gestión de datos e incertidumbre	20	19,6 %
Total	102	100 %

Tabla D.2 Matemática ENLA 2024. Distribución de ítems por capacidades en la prueba

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

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.	7	6,9 %
	Usa estrategias y procedimientos para recopilar y procesar datos.	3	2,9 %
	Sustenta conclusiones o decisiones con base en la información obtenida.	9	8,8 %
Total		102	100 %

El total ha sido redondeado a un decimal. Por ello, la suma total en esta tabla no es exactamente 100 %.

Tabla D.3 Matemática ENLA 2024. Distribución de ítems por contextos en la prueba

Contextos	Cantidad de ítems	Porcentaje
Intramatemático	16	15,7%
Extramatemático	86	84,3%
Total	102	100,0%

Ministerio de Educación

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

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