

COMPETENCIA
LOGROS
LECTURA
MATEMÁTICA
MEDICIÓN
MUESTREO
DESEMPEÑO
PSICOMETRÍA
APRENDIZAJE
EDUCACIÓN
CALIDAD

Reporte técnico de las Evaluaciones Censales y Muestrales de Estudiantes 2019

COMPETENCIA
LOGROS
LECTURA
MATEMÁTICA
MEDICIÓN
MUESTREO
DESEMPEÑO
PSICOMETRÍA
APRENDIZAJE
EDUCACIÓN
CALIDAD

Reporte técnico de las Evaluaciones Censales y Muestrales de Estudiantes 2019



PERÚ

Ministerio
de Educación

Carlos Martín Benavides Abanto

Ministro de Educación del Perú

Sandro Luis Parodi Sifuentes

Viceministro de Gestión Institucional

Diana Mariela Marchena Palacios

Viceministra de Gestión Pedagógica

Gabriela María Carrasco Carrasco

Secretario de Planificación Estratégica

Humberto Pérez León Ibáñez

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

Reporte técnico de las Evaluaciones Censales y Muestrales de Estudiantes 2019

Responsables del documento

Andrés Burga León (coordinador)

Yuriko Sosa Paredes

Luis Mejía Campos

Tania Pacheco Valenzuela

Yoni Arámbulo Mogollón

Gabriela Santibáñez Rojas

Wílmer Hernández Cabrera

Ángel Zegarra López

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, 2020

Calle Del Comercio 193, San Borja

Lima, Perú

Teléfono: (511) 615-5800

www.minedu.gob.pe

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

En el presente documento, se utilizan de manera inclusiva términos como “el docente”, “el estudiante” y sus respectivos plurales (así como otras palabras equivalentes en el contexto educativo) para referirse a hombres y mujeres. Esta opción se basa en una convención idiomática y tiene por objetivo evitar las formas para aludir a ambos géneros en el idioma castellano (“o/a”, “los/las” y otras similares), debido a que implican una saturación gráfica que puede dificultar la comprensión lectora.

Índice

Introducción	6
Capítulo 1: Construcción de las pruebas aplicadas en la Evaluación Censal de Estudiantes (ECE)	7
1.1 Construcción de las pruebas aplicadas en la ECE	8
1.1.1 Construcción de los ítems y las pruebas	8
1.1.2 Procesos de validación de los instrumentos de evaluación	11
1.1.2.1 Validación de expertos de la UMC	11
1.1.2.2 Validación de campo: aplicación piloto	11
1.1.2.3 Análisis de las propiedades psicométricas y de la distribución de los ítems en la escala	12
1.1.2.4 Validación de expertos de otras oficinas del Minedu	13
1.1.2.5 Adaptación de instrumentos para estudiantes con discapacidad	13
1.1.2.6 Aplicación definitiva	15
1.1.3 El ensamblaje de las pruebas	15
Capítulo 2: Población y muestra	27
2.1 Población objetivo	28
2.2 Marco poblacional	28
2.3 Cobertura	29
2.4 Ajuste por no respuesta y cálculo de resultados	30
Capítulo 3: Operativo de campo	37
3.1 Organización territorial del operativo	39
3.2 Conformación de la red administrativa de aplicación	41
3.3 Proceso de selección y capacitación del personal de la RA	43
3.4 Características de los instrumentos y documentos utilizados	46
3.5 Proceso de aplicación	48
Capítulo 4: Procesamiento de datos y análisis psicométrico	53
4.1 Gestión y depuración de datos	54
4.1.1 Gestión física	54
4.1.2 Captura de datos	55
4.1.3 Depuración de datos	55
4.2 Procesamiento psicométrico	56
4.2.1 Modelo Rasch para ítems dicotómicos	56
4.3 Modelo Rasch de crédito parcial	59
4.4 Análisis de ítems	63
4.4.1 Calibración de ítems	65

4.4.2	Evidencias de validez vinculadas a la estructura interna de las pruebas	79
4.4.3	Confiabilidad y consistencia de la clasificación	81
4.4.4	Equiparación de medidas	84
4.5	Niveles de logro y presentación de resultados	91
	Referencias	95

Introducción

La Oficina de Medición de la Calidad de los Aprendizajes (UMC) de la Secretaría de Planificación Estratégica del Ministerio de Educación (Minedu) tiene entre sus funciones el diseño y la implementación de las evaluaciones de logros de aprendizaje de los estudiantes de Educación Básica a nivel nacional. De esta manera, se busca producir información referida al logro de los aprendizajes mínimos esperados según el currículo oficial, que pueda ser utilizada para mejorar los aprendizajes en las áreas y competencias evaluadas tanto por el sistema escolar como por sus actores, entre ellos los padres de familia, directores y docentes de todas las escuelas, especialistas y autoridades de las Unidades de Gestión Educativa Local (UGEL), Direcciones Regionales de Educación (DRE) y Minedu.

En ese marco, de acuerdo a la RM 168-2019-Minedu, la UMC llevó a cabo en el 2019 la Evaluación Censal de Estudiantes (ECE) y la Evaluación Muestral (EM). La ECE evaluó, en 2.º grado de secundaria, las áreas de Comunicación (Lee diversos tipos de textos escritos en su lengua materna), Matemática (todas las competencias) y Ciencia y Tecnología (todas las competencias). Asimismo, la EM evaluó en 2.º y 4.º grado de primaria las áreas de Comunicación (Lee diversos tipos de textos escritos en su lengua materna) y Matemática (todas las competencias).

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 muestrales y los pesos utilizados en la estimación de los diversos parámetros. Este capítulo incluye además el análisis de la cobertura.

El tercer capítulo aborda los aspectos relacionados con el operativo de campo. Describe cómo se organizó la red administrativa, cómo fueron seleccionados y capacitados los aplicadores, cuáles fueron los instrumentos utilizados y cómo se recolectaron los datos en el contexto de las evaluaciones 2019.

Finalmente, 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. Además, se describe cómo se realizó el proceso de equiparación de medidas y la forma en que se establecieron los niveles de logro a partir de los puntos de corte.

Construcción de las pruebas aplicadas en la Evaluación Censal de Estudiantes (ECE)

Capítulo 1

1.1 Construcción de las pruebas aplicadas en la ECE

La UMC cuenta con un equipo de evaluación conformado por especialistas de Matemática, Comunicación, Educación Intercultural Bilingüe, Ciencias Sociales y Ciudadanía, Ciencia y Tecnología. Este equipo es el encargado de diseñar y construir las pruebas y demás instrumentos para las diversas evaluaciones y los estudios que están bajo la responsabilidad de la UMC. Además, el equipo de Atención a la Diversidad orienta y contribuye a la adaptación de los diversos instrumentos a las características particulares de los estudiantes con discapacidad incluidos en las escuelas de Educación Básica regular, con la finalidad de que participen en las evaluaciones.

1.1.1 Construcción de los ítems y las pruebas

El diseño y la construcción tanto de los ítems como de las pruebas siguen los principios de validez, confiabilidad y diseño universal de evaluación. Estos establecen 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). Asimismo, los instrumentos deben reflejar una concepción inclusiva de la educación, conforme a los lineamientos de la política educativa nacional (Ministerio de Educación, 2015).

En la ECE y en la EM, se emplean ítems de dos formatos: de opción múltiple y de respuesta construida.

Ítem de opción múltiple. El formato de opción múltiple (el más usual) presenta en su estructura un estímulo, enunciado o pregunta que se completa con tres o cuatro opciones 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 algunos ítems, alguno de los distractores puede constituir una respuesta intermedia o de logro parcial y, por lo tanto, esta se modela como un ítem de crédito parcial (puntaje parcial). En las evaluaciones implementadas por la UMC, las pruebas de 2.º grado de primaria y 4.º grado de primaria EIB en su mayoría están conformadas por ítems que presentan tres alternativas para evitar que el estudiante, debido a su edad o al aprendizaje de la

lectoescritura en contexto EIB, respectivamente, haga un mayor esfuerzo cognitivo en la lectura que en la solución del ítem. Para los demás grados, las pruebas están conformadas por ítems que presentan cuatro opciones.

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

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.
- El ítem debe estar libre de todo sesgo sociocultural, socioeconómico, de género o de religión.
- El ítem no debe estar relacionado con ningún otro ítem de la prueba; es decir, 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, el tamaño y el tipo de fuente, los espacios entre caracteres y entre líneas, el contraste, etc., 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, 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 opción correcta debe ser solo una y debe estar acompañada por distractores plausibles.
- A lo largo de la prueba, la opción correcta debe estar distribuida entre las distintas ubicaciones.
- Las opciones deben ser preferiblemente tres o cuatro.
- Las opciones deben disponerse, de preferencia, de forma vertical.
- Las alternativas de cada ítem debe organizarse de acuerdo con un criterio; por ejemplo, del número menor al número mayor.
- Las opciones deben ser autónomas entre sí, sin solaparse ni referirse unas a otras.
- Se deben evitar en todo momento las opciones “Todas las anteriores” y “Ninguna de las anteriores”.
- Ninguna opción 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 Lectura) que le da un marco tanto al ítem como a la respuesta que se espera. Este tipo de ítem también puede incluir instrucciones o especificaciones para responder.

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 involucrados 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 y 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, los tipos de respuesta (de crédito total, de crédito parcial y sin crédito) y establece criterios de codificación para cada uno de estos tipos de respuesta. Además, este 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. Un estudiante puede utilizar diferentes procedimientos para resolver la tarea. 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 tanto, no responden a los criterios de codificación establecidos.

En las pruebas de la ECE, los ítems de respuesta construida se encuentran en menor cantidad que los ítems de opción múltiple, porque el procesamiento de las respuestas a este tipo de ítems es complejo y, en aplicaciones masivas, demanda muchos recursos y tiempo. En cambio, en las pruebas de la EM, debido al menor volumen de respuestas a codificar, se incluye una mayor cantidad de ítems de respuesta construida.

1.1.2 Procesos de validación de los instrumentos de evaluación

1.1.2.1 Validación de expertos de la UMC

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

1.1.2.2 Validación de campo: aplicación piloto

Después de la revisión de expertos al interior de la UMC, los equipos 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. Esta 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 forman 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 estructura 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 en las diversas 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 que se liberan; es decir, que son publicados en los informes de resultados que produce y distribuye la UMC.

1.1.2.3 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 de los ítems; además, se realiza una revisión pedagógica de la distribución de los ítems en la escala. Entre los aspectos psicométricos, se analiza el ajuste de los ítems y de los distractores, la confiabilidad de las medidas, la unidimensionalidad, la correspondencia entre la media de habilidad de los estudiantes y la media de dificultad de cada prueba, y el comportamiento diferencial de los ítems según los distintos estratos (urbano y rural, estatal y no estatal, hombre y mujer). Igualmente, se ensayan hipótesis sobre las posibles causas por las que un ítem no ajustó al modelo psicométrico, tuvo un funcionamiento diferencial según los estratos o su dificultad no concordó con la complejidad esperada. Para una mayor profundización de este aspecto, en el capítulo 4 de este documento se detallan dichas propiedades psicométricas.

En la revisión pedagógica, se analiza la correspondencia entre la complejidad pedagógica de un ítem y la dificultad dada por el modelo psicométrico, 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 de Comunicación, tanto en primaria como en secundaria, también se analizan los

textos para encontrar posibles explicaciones a las respuestas no esperadas, a un error recurrente de los estudiantes o al comportamiento diferencial de algunos ítems.

El proceso de análisis concluye con la selección de los ítems de mejor ajuste y desempeño para la aplicación definitiva de las pruebas.

1.1.2.4 Validación de expertos de otras oficinas del Minedu

Según lo establecido por la UMC, el proceso de validación incluye que un conjunto de expertos en cada una de las áreas evaluadas emita su juicio (criterio de jueces) respecto de la construcción, la correspondencia con el indicador y la dificultad de la tarea solicitada en cada ítem empleado para medir el constructo implicado. Para ello, los ítems seleccionados a partir de la aplicación piloto y el análisis posterior son organizados en formas de prueba, para que sean validados por expertos de las distintas direcciones de línea del Minedu tales como la Dirección de Educación Primaria (DEP), la Dirección de Educación Secundaria (DES), la Dirección General de Educación Básica Alternativa, Intercultural Bilingüe y de Servicios Educativos (Digeibira)¹. En el proceso de validación, también participan especialistas de la Dirección de Evaluación Docente (DIED) por su experticia en la construcción de ítems y su nivel de aporte. En esta validación, se solicita a los jueces que evalúen los aspectos mencionados usando una ficha técnica.

Esta ficha 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 según las características de la competencia que evalúan. 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, estime convenientes. Luego, las fichas son sistematizadas, los cuestionamientos absueltos y las sugerencias incorporadas. Ello implica, en algunos casos, realizar algún tipo de ajuste al ítem o al texto.

1.1.2.5 Adaptación de instrumentos para estudiantes con discapacidad

Desde el 2015, la UMC adapta (es decir, realiza cambios o ajustes razonables) los instrumentos de evaluación de las distintas competencias evaluadas, así como los procedimientos de aplicación de las pruebas, con el fin de asegurar la participación plena de los estudiantes con discapacidad sensorial (baja visión, ceguera, hipoacusia y sordera), discapacidad motora y trastorno del espectro autista en las

¹La prueba de 4.º grado de primaria EIB, además, es revisada por especialistas de UGEL de las regiones correspondientes a las lenguas originarias evaluadas, así como docentes de aula. Asimismo, los ítems de esta prueba son sometidos a un proceso de doble traducción (con expertos traductores propios de las culturas originarias) con la finalidad de que sean pertinentes a las características socioculturales y lingüísticas de los estudiantes evaluados.

pruebas nacionales a gran escala. 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 o procedimientos de evaluación 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 como a los demás estudiantes.

En 2019, se implementaron las siguientes acomodaciones en las pruebas y procesos de aplicación:

Tabla 1.1 Acomodaciones a las pruebas de la ECE y la EM 2019

Tipo de discapacidad	Tipo de acomodación	Descripción		
Visual	Presentación	Prueba en macrotipo		
	Baja visión	Equipos y materiales	Plumón delgado, lupa, mica amarilla, tiposcopio	
	Horario y tiempo		Menos ítems que en la prueba regular	
	Ubicación		Carpeta junto a acceso de luz natural	
	Ceguera total	Presentación	Prueba en braille e imágenes en alto	
		Equipos y materiales	Regleta y punzón, ábaco, cubarritmo, reglas en braille	
		Horario y tiempo		Menos ítems que en la prueba regular
		Ubicación		Carpeta al fondo del aula
	Respuesta		<i>Stickers</i> para marcar respuestas en el cuadernillo	
Auditiva	Hipoacusia / Sordera	Presentación	Léxico y sintaxis simplificados	
		Horario y tiempo		Menos ítems que en la prueba regular
		Ubicación		Carpeta al fondo del aula
		Respuesta		Intérprete de lengua de señas peruana
Sordoceguera			Según la severidad, lo mismo que para baja visión o ceguera	
Física	Presentación		Macrotipo (si lo requiere)	
	Equipos y materiales		Lápiz jumbo (si lo requiere; por ejemplo, para estudiantes que escriben con los pies)	
	Respuesta		Persona de apoyo para marcar y pasar las hojas	
Trastorno del espectro autista	Autismo	Presentación		Léxico y sintaxis simplificados
		Equipos y materiales		Tarjetas de comunicación alternativa
		Horario y tiempo		Menos ítems que en la prueba regular
		Ubicación		Carpeta al fondo del aula
				Pautas personalizadas por parte del aplicador especializado

1.1.2.6 Aplicación definitiva

La aplicación definitiva de la ECE y la EM se realizó en el último trimestre del año (octubre-diciembre). En el caso de la ECE, la aplicación se llevó a cabo en 2.º de secundaria durante tres días. En el primer día, primero se aplicó la prueba de Lectura y luego la de Matemática. En el segundo día, el orden de aplicación de las pruebas fue inverso. Este cambio obedece a la necesidad de neutralizar los posibles efectos del cansancio de los estudiantes, quienes rinden dos pruebas en un día. El tercer día se destinó a la aplicación de la prueba de Ciencia y Tecnología. Por su parte, la EM de 2.º y de 4.º grado de primaria se realizó en dos días. El primer día, primero se aplicó la prueba de Lectura y luego la de Matemática. En el segundo día, el orden de aplicación de las pruebas fue inverso.

A partir del 2015, los instrumentos de la ECE, incluidos los cuadernillos de las pruebas, retornan en su totalidad al Minedu y se mantienen bajo confidencialidad. Este procedimiento, aunque complicado, se realiza porque permite alimentar un banco de ítems creciente y robustecer los procesos de equiparación, y, por ende, la comparabilidad de resultados en el tiempo.

En resumen, mediante los procesos de construcción y validación, se logró que los instrumentos de evaluación de la ECE y de la EM 2019 cumplan con los requerimientos técnicos y las consideraciones de contenido que permiten recoger datos válidos y confiables sobre los logros de aprendizaje de los estudiantes evaluados. Asimismo, se aseguró que los instrumentos guarden correspondencia con el currículo y otros documentos pedagógicos del sistema educativo, y que sus resultados sean comparables a lo largo del tiempo.

1.1.3 El ensamblaje de las pruebas

Conforme avanzan en su escolaridad, los estudiantes amplían y profundizan los aprendizajes logrados en los grados anteriores. Así, por ejemplo, respecto de la Lectura, el lenguaje de los textos se complejiza y se especializa progresivamente, los temas tratados son más diversos, y las capacidades de inferencia y reflexión se consolidan. 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 la manera más precisa lo que realmente logran hacer los estudiantes. Sin embargo, las evaluaciones de carácter censal suelen enfrentar la restricción de tener que medir con una prueba única (y las consiguientes limitaciones del tiempo de evaluación y del número de ítems) las habilidades de los estudiantes. Para lidiar con esta restricción, en la evaluación, se estudió la posibilidad de utilizar diseños que contengan bloques de ítems comunes, que sean aplicados por todos los

estudiantes, y bloques no comunes, que sean aplicados solo por un conjunto de estudiantes. La combinación de bloques comunes y no comunes debía permitir el ensamblaje de formas (o “cuadernillos”) que se distribuyeran aleatoriamente entre la población. De este modo, y gracias al uso de técnicas de equiparación del modelo Rasch, una prueba podría contar con mayor cantidad de ítems en la misma métrica, sin que ello implicara incrementar la cantidad de ítems a los que cada estudiante se enfrenta.

Con la asistencia técnica del Berkeley Evaluation and Assessment Research Center (BEAR), de la Universidad de California (Berkeley), se llevaron a cabo estudios psicométricos que plantearon diversos escenarios en los cuales se combinaban bloques de ítems de maneras distintas para ensamblar las formas de las pruebas, a fin de ampliar el número total de ítems aplicados (aunque no todos los estudiantes respondieran el total de estos ítems). En estos escenarios se consideraban bloques comunes, bloques no comunes y un número determinado de ítems por bloque. A partir de estos escenarios, se realizaron simulaciones estadísticas con la base de datos de la aplicación piloto, para verificar que la confiabilidad de las personas en estos diseños se mantuviera estable en múltiples iteraciones, de manera que fuera factible, dentro de márgenes razonables, reportar resultados a nivel individual y/o de agregado por escuela. En todos los escenarios planteados, se obtuvieron niveles altos en la media de la confiabilidad (entre 0,90 y 0,96), así como en la media de la correlación entre el valor verdadero del rasgo latente y su valor estimado (entre 0,95 y 0,96).

En relación con la conformación de los bloques, es preciso señalar que el bloque común tiene por finalidad optimizar la equiparación de las medidas y es construido de tal forma que los ítems reflejen la tabla de especificaciones, así como la dispersión de las medidas a lo largo de la escala de dificultad.

Por otro lado, en una evaluación muestral es posible utilizar diseños más sencillos, como el diseño de bloques balanceados incompletos (BIBD, por sus siglas en inglés), pues no se tiene las exigencias que supone una devolución de resultados a nivel de estudiante. Esta ventaja ha sido aprovechada para el diseño de las pruebas de Lectura y Matemática de 2.º grado de primaria.

Matrices de bloques de las pruebas de la ECE 2019

Matriz de bloques de la prueba de Lectura de 2.º grado de secundaria

La prueba de Lectura de 2.º grado de secundaria elaborada para la ECE 2019 cuenta con un total de 16 textos de diferente tipo, género, formato y complejidad. A partir de este conjunto de textos, se elaboraron 98 ítems. Dichos ítems se distribuyeron en 2 bloques comunes de 19 ítems cada uno (un bloque común por cada día de aplicación) y 10 bloques no comunes de 6 ítems cada uno. Las diferentes combinaciones de los

bloques permitieron ensamblar 10 formas que reflejan los indicadores de la tabla de especificaciones y presentan una dificultad similar entre sí.

Tabla 1.2 *Matriz de bloques de la prueba de Lectura. Día 1*

Forma	Bloque	Ítems por forma
1	B01	25
	B03	
2	B01	25
	B04	
3	B01	25
	B05	
4	B01	25
	B06	
5	B01	25
	B07	

Tabla 1.3 *Matriz de bloques de la prueba de Lectura. Día 2*

Forma	Bloque	Ítems por forma
6	B02	25
	B08	
7	B02	25
	B09	
8	B02	25
	B10	
9	B02	25
	B11	
10	B02	25
	B12	

De las 10 formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 98 ítems que tiene la prueba, cada estudiante respondió 50 (25 por día).

Matriz de bloques de la prueba de Matemática de 2.º grado de secundaria

La prueba de Matemática de 2.º grado de secundaria elaborada para la ECE 2019 cuenta con un total de 90 ítems distribuidos en 10 formas. Cada forma contiene 25 ítems: 20 ítems comunes organizados en 6 bloques comunes y 5 ítems no comunes que corresponden a un bloque que varía de una forma a otra. Las diferentes combinaciones de los bloques permitieron organizar formas que reflejan los

indicadores de la tabla de especificaciones y presentan una dificultad similar entre sí.

Tabla 1.4 Matriz de bloques de la prueba de Matemática. Día 1

Forma	Bloque	Ítems por forma
1	B01	25
	B11	
	B19	
	B12	
	B20	
	B13	
	B14	
2	B02	25
	B13	
	B19	
	B14	
	B20	
	B11	
	B12	
3	B03	25
	B11	
	B19	
	B12	
	B20	
	B13	
	B14	
4	B04	25
	B13	
	B19	
	B14	
	B20	
	B11	
	B12	
5	B05	25
	B11	
	B19	
	B12	
	B20	
	B13	
	B14	

Tabla 1.5 Matriz de bloques de la prueba de Matemática. Día 2

Forma	Bloque	Ítems por forma
6	B06	25
	B15	
	B21	
	B16	
	B22	
	B17	
7	B18	25
	B07	
	B17	
	B21	
	B22	
	B15	
8	B16	25
	B08	
	B15	
	B21	
	B22	
	B17	
9	B18	25
	B09	
	B17	
	B21	
	B22	
	B15	
10	B16	25
	B10	
	B15	
	B21	
	B22	
	B17	
	B18	

De las 10 formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 90 ítems que tiene la prueba, cada estudiante respondió 50 (25 por día).

Matriz de bloques de la prueba de Ciencia y Tecnología de 2.º grado de secundaria

La prueba de Ciencia y Tecnología de 2.º grado de secundaria elaborada para la ECE 2019 cuenta con un total de 94 ítems distribuidos en 6 bloques de 16 ítems cada uno. Las diferentes combinaciones de los bloques permitieron ensamblar 6 formas que reflejan los indicadores de la tabla de especificaciones y presentan una dificultad similar entre sí.

Tabla 1.6 Matriz de bloques de la prueba de Ciencia y Tecnología. Día 3

Forma	Bloque	Ítems por bloque	Ítems por forma
1	B01	16	32
	B02	16	
2	B03	16	32
	B04	16	
3	B05	16	32
	B06	16	
4	B02	16	32
	B03	16	
5	B04	16	32
	B05	16	
6	B06	16	32
	B01	16	

De las 6 formas ensambladas, cada estudiante evaluado se enfrentó a una de ellas. En consecuencia, de los 94 ítems que tiene la prueba, cada estudiante respondió 32.

Matrices de bloques de las pruebas de la EM 2019

Matriz de bloques de la prueba de Lectura de 2.º grado de primaria

La prueba de Lectura de 2.º grado de primaria elaborada para la EM 2019 cuenta con un total de 21 textos de diferente tipo, género, formato y complejidad. A partir de este conjunto de textos, se elaboraron 95 ítems. Dichos ítems se distribuyeron en 2 bloques comunes y 22 bloques no comunes. El bloque común 01 (B01) está conformado por 14 ítems y el otro (B02), por 10 ítems. Los bloques no comunes están conformados por 6 ítems, 5 ítems o 1 ítem. Las diferentes combinaciones de los bloques permitieron ensamblar 8 formas que reflejan los indicadores de la tabla de especificaciones y presentan una dificultad diferenciada según el área de ubicación de las escuelas en que se aplicaron. Así, de las 8 formas, 4 tenían una mayor dificultad y se aplicaron

a los estudiantes de escuelas de las áreas urbanas de las diferentes regiones. Las otras 4 tenían una menor dificultad y se aplicaron a los estudiantes de escuelas de las áreas rurales y a la totalidad de los estudiantes de 2.º grado de primaria de Loreto y Ucayali.

Tabla 1.7 Matriz de bloques de la prueba de Lectura. Día 1

Forma	Bloque	Ítems por forma
1	B01	25
	B03	
	B04	
2	B01	25
	B05	
	B06	
3	B07	25
	B01	
	B08	
4	B09	25
	B10	
	B01	
	B11	
	B04	

Tabla 1.8 Matriz de bloques de la prueba de Lectura. Día 2

Forma	Bloque	Ítems por forma
5	B02	25
	B12	
	B13	
	B14	
6	B02	25
	B12	
	B15	
7	B16	25
	B17	
	B02	
	B18	
	B19	
8	B16	25
	B20	
	B02	
	B18	
	B21	
	B22	

De las 8 formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 95 ítems que tiene la prueba, cada estudiante respondió 50 (25 por día).

Matriz de bloques de la prueba de Matemática de 2.º grado de primaria

La prueba de Matemática de 2.º grado de primaria elaborada para la EM 2019 cuenta con un total de 88 ítems distribuidos en 8 formas. Cada forma contiene 23 ítems: 16 ítems comunes organizados en 2 bloques comunes y 7 ítems no comunes que corresponden a un bloque que varía de una forma a otra. Las diferentes combinaciones de los bloques permitieron organizar formas que reflejan los indicadores de la tabla de especificaciones y presentan una dificultad diferenciada, según el área de ubicación de las escuelas en que se aplicaron. Así, de las 8 formas, 4 tenían una mayor dificultad y se aplicaron a los estudiantes de escuelas de las áreas urbanas de las diferentes regiones. Las otras 4 tenían una menor dificultad y se aplicaron a los estudiantes de escuelas de las áreas rurales y a la totalidad de los estudiantes de 2.º grado de primaria de Loreto y Ucayali.

Tabla 1.9 Matriz de bloques de la prueba de Matemática. Día 1

Forma	Bloque	Ítems por forma
1	B09	23
	B01	
	B10	
2	B10	23
	B02	
	B09	
3	B09	23
	B03	
	B10	
4	B10	23
	B04	
	B09	

Tabla 1.10 Matriz de bloques de la prueba de Matemática. Día 2

Forma	Bloque	Ítems por forma
5	B11	23
	B05	
	B12	
6	B12	23
	B06	
	B11	
7	B11	23
	B07	
	B12	
8	B12	23
	B08	
	B11	

De las 8 formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 88 ítems que tiene la prueba, cada estudiante respondió 46 (23 por día).

Matriz de bloques de la prueba de Lectura de 4.º grado de primaria

La prueba de Lectura de 4.º grado de primaria elaborada para la EM 2019 cuenta con un total de 12 textos de diferentes tipo, género, formato y complejidad. A partir de este conjunto de textos, se elaboraron 72 ítems. Dichos ítems se distribuyeron en 2 bloques comunes (uno por día de aplicación) de 18 ítems cada uno y 6 bloques no comunes de 6 ítems cada uno. Las diferentes combinaciones de los bloques permitieron ensamblar

6 formas que reflejan los indicadores de la tabla de especificaciones y presentan una dificultad similar entre sí.

Tabla 1.11 Matriz de bloques de la prueba de Lectura. Día 1

Forma	Bloque	Ítems por forma
1	B01	24
	B02	
2	B01	24
	B03	
3	B01	24
	B04	

Tabla 1.12 Matriz de bloques de la prueba de Lectura. Día 2

Forma	Bloque	Ítems por forma
4	B05	24
	B06	
5	B05	24
	B07	
6	B05	24
	B08	

De las 6 formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 72 ítems que tiene la prueba, cada estudiante respondió 48 (24 por día).

Matriz de bloques de la prueba de Matemática de 4.º grado de primaria

La prueba de Matemática de 4.º grado de primaria elaborada para la ECE 2019 cuenta con un total de 70 ítems distribuidos en 6 formas. Cada forma contiene 25 ítems: 20 ítems comunes organizados en 6 bloques comunes y 5 ítems no comunes que corresponden a un bloque que varía de una forma a otra. Las diferentes combinaciones de los bloques permitieron organizar formas que reflejan los indicadores de la tabla de especificaciones y presentan una dificultad similar entre sí.

Tabla 1.13 Matriz de bloques de la prueba de Matemática. Día 1

Forma	Bloque	Ítems por forma
1	B01	25
	B07	
	B15	
	B08	
	B16	
	B09	
	B10	
2	B02	25
	B09	
	B15	
	B10	
	B16	
	B07	
3	B03	25
	B07	
	B15	
	B08	
	B16	
	B09	
	B10	

Tabla 1.14 Matriz de bloques de la prueba de Matemática. Día 2

Forma	Bloque	Ítems por forma
4	B04	25
	B11	
	B17	
	B12	
	B18	
	B13	
	B14	
5	B05	25
	B11	
	B17	
	B12	
	B18	
	B13	
	B14	
6	B06	25
	B13	
	B17	
	B14	
	B18	
	B11	
B12		

De las 6 formas ensambladas, cada estudiante evaluado se enfrentó a dos de ellas (una por cada día de aplicación). En consecuencia, de los 70 ítems que tiene la prueba, cada estudiante respondió 50 (25 por día).

Población y muestra

Capítulo 2

2.1 Población objetivo

La ECE 2019 tuvo como población objetivo a los estudiantes que asistieron a clases en el 2.º grado de secundaria de una IE de Educación Básica Regular (EBR).

Se evaluó a los estudiantes de 2.º grado de secundaria porque en este se concluye el 6.º ciclo de la EBR. Esto 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 una IE que atiende a menos de cinco estudiantes matriculados ² en el grado a ser evaluado. En secundaria, la exclusión fue del 0,6 % de los estudiantes a nivel nacional; las razones de esta exclusión se encuentran detalladas en el Marco de trabajo de la ECE (Ministerio de Educación, 2009).

En el 2019, la ECE tenía programado evaluar 13 441 IE, 23 989 secciones y aproximadamente 527 498 estudiantes en 2.º grado de secundaria.

2.2 Marco poblacional

El marco poblacional ha sido elaborado 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 2018 y la información del Sistema de Información de Apoyo a la Gestión de la Institución Educativa (Siagie) del 2019. 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 sobre la cantidad de estudiantes y secciones que proviene del Censo Educativo es declarada por el Director de la IE. En el Censo Educativo 2018, existen algunas IE que no reportan datos. Por ello, la información corresponde a datos imputados. El porcentaje de escuelas con datos de matrícula imputados en el Censo Escolar 2018 es alrededor del 1,5 %.

Para elaborar el listado de IE a evaluar en la ECE 2019, se aplicó el siguiente procedimiento:

²La 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 o de la aplicación de la ECE del año anterior.

1. Se utilizó el Padrón de Instituciones Educativas para determinar el listado de IE activas³. Se seleccionaron las IE de secundaria de EBR.
2. Se calculó la cantidad de secciones y estudiantes programados. Esta información corresponde al valor registrado en el Siagie en caso las nóminas de matrículas se encuentren aprobadas al momento de la construcción del marco poblacional. Si una IE no contó con las nóminas de matrícula aprobadas, se tomó el valor mayor entre la información del Censo Educativo 2018, del Siagie 2019 o de la cantidad de estudiantes evaluados en la ECE 2018.
3. Se seleccionaron las IE que tuvieron cinco o más estudiantes asignados en el punto 2.

2.3 Cobertura

La cobertura a nivel de estudiantes se muestra en la tabla 2.1. El porcentaje de estudiantes evaluados respecto de los programados a nivel nacional es del 97 %. 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 2019.

Tabla 2.1 *Porcentaje de estudiantes evaluados por estratos a nivel nacional y en cada DRE*

		Estatal	No estatal	Urbana	Rural	Total
	Nacional	97,3	96,1	97,5	93,3	97,0
0100	Amazonas	95,3	87,4	97,7	90,0	95,2
0200	Áncash	98,2	96,0	97,9	97,5	97,8
0300	Apurímac	98,1	96,8	98,3	97,2	98,1
0400	Arequipa	99,1	98,1	98,8	96,0	98,8
0500	Ayacucho	98,6	97,3	98,8	97,3	98,5
0600	Cajamarca	97,9	94,8	98,5	95,7	97,6
0701	Callao	98,2	95,2	97,1	-	97,1
0800	Cusco	98,6	97,4	98,6	97,8	98,4
0900	Huancavelica	96,4	97,4	96,9	95,6	96,4
1000	Huánuco	96,2	96,3	97,5	92,2	96,2
1100	Ica	98,4	98,6	98,5	97,1	98,5
1200	Junín	97,0	95,4	97,4	91,5	96,7
1300	La Libertad	96,9	96,2	97,4	92,7	96,8
1400	Lambayeque	97,2	94,8	96,6	94,3	96,5
1501	Lima Metropolitana	98,2	95,7	97,1	98,1	97,1
1502	Lima Provincias	98,1	97,1	97,9	97,2	97,9
1600	Loreto	90,3	96,1	93,8	79,2	90,6
1700	Madre de Dios	97,9	93,3	97,4	98,5	97,6

³El 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/>

		Estatal	No estatal	Urbana	Rural	Total
1800	Moquegua	99,6	99,6	99,6	100,0	99,6
1900	Pasco	95,0	96,0	98,3	86,3	95,1
2000	Piura	98,3	96,8	98,2	97,0	98,0
2100	Puno	98,9	98,7	99,0	98,5	98,9
2200	San Martín	96,4	97,3	97,6	90,4	96,4
2300	Tacna	99,4	99,3	99,4	99,3	99,4
2400	Tumbes	97,0	98,0	97,1	100	97,2
2500	Ucayali	92,5	97,4	95,1	80,1	92,8

2.4 Ajuste por no respuesta y cálculo de resultados

El ajuste por no respuesta es un ponderador utilizado para expandir la población evaluada, con el fin de tener en cuenta a los estudiantes no evaluados. Este ajuste es realizado a nivel de estrato, el cual es construido a partir de la combinación entre DRE, gestión, área y característica. Para el cálculo de resultados en 2.º grado de secundaria, es necesario multiplicar los datos de cada estudiante por el ajuste por no respuesta. Los resultados se reportan según los siguientes niveles:

- nacional,
- por sexo de los estudiantes (femenino, masculino),
- por gestión de la IE (estatal, no estatal),
- por área geográfica de la IE (urbana, rural),
- por Dirección Regional de Educación (DRE),
- por Unidad de Gestión Educativa Local (UGEL),
- por distrito,

El cálculo del ajuste por no respuesta se realizó de la siguiente manera:

Paso 1

$$aj_{IE_i} = \frac{p_{IE_i}}{e_{IE_i}}$$

donde

e_{IE_i} = Cantidad de alumnos evaluados en la i-ésima IE.

p_{IE_i} = Cantidad de alumnos programados en la i-ésima IE.

aj_{IE_i} = Ajuste por no respuesta en la i-ésima IE.

Paso 2

$$aj_{E_j} = \frac{p_{E_j}}{e_{E_j}}$$

donde

e_{E_j} = Cantidad de alumnos programados en el j-ésimo estrato.

p_{E_j} = Cantidad de alumnos evaluados ajustada en el j-ésimo estrato, calculado ponderando los alumnos evaluados por el ajuste por no respuesta en la IE.

aj_{E_j} = Ajuste por no respuesta en el j-ésimo estrato.

Finalmente, se calculó el ajuste por no respuesta para cada estudiante.

$$aj_{NR} = aj_{IE_i} \times aj_{E_j}$$

Evaluaciones muestrales

Dentro de las evaluaciones de rendimiento del 2019 se llevaron a cabo dos evaluaciones muestrales. Específicamente, se evaluó Lectura y Matemática en 2.º y en 4.º grado de primaria. La construcción del marco muestral se llevó a cabo siguiendo el mismo procedimiento que en las evaluaciones censales; es decir, excluyendo a aquellas IE en las que asistieron menos de cinco estudiantes en el grado a ser evaluado.

El muestreo empleado fue de tipo probabilístico, estratificado, por conglomerados y de dos etapas, donde la unidad primaria de muestreo corresponde a las IE y la unidad secundaria, 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. En la segunda etapa, se realizó una selección aleatoria simple de hasta dos secciones.

Conglomerados

Cada IE es considerada un conglomerado, pues agrupa a un conjunto de secciones que, a su vez, agrupan a un conjunto de estudiantes. El muestreo de conglomerados surge como una necesidad práctica, económica y de eficiencia administrativa. Su mayor ventaja es que no se necesita contar con un marco muestral de cada elemento por analizar; solo es necesario tener el marco muestral a nivel de conglomerados (IE y secciones) para realizar la selección de los mismos (Lohr, 1999). 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 2019, se realizó una estratificación principalmente porque, en un contexto tan diverso como el peruano, la variable de interés (rendimiento) toma diferentes valores en distintas subpoblaciones definidas por sus diferencias según el área geográfica, gestión o característica⁴ de la IE. 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 primaria, se conformaron los siguientes estratos:

- Estrato 1: Urbano - Estatal - Polidocente completo
- Estrato 2: Urbano - Estatal - Unidocente / Multigrado
- Estrato 3: Rural - Estatal - Polidocente completo
- Estrato 4: Rural - Estatal - Unidocente / Multigrado
- Estrato 5: No estatal (las IE no estatales, casi en su totalidad, son urbanas y polidocentes)

La evaluación muestral considera además 26 dominios que corresponden a cada DRE, pues estas instancias ejercen control administrativo en diferentes conjuntos de IE. Por ello, las DRE son combinadas con los estratos antes mencionados, salvo en el caso de Lima Metropolitana y Callao, que solo consideran los estratos 1 y 2. Por ello, el total de estratos considerados es de 124.

El tamaño de la muestra es determinado por cada dominio y, una vez estratificado, 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 de los mismos.

En las tablas 2.2 y 2.3, 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.

⁴Solo aplicable para primaria.

Tabla 2.2 Cobertura de la EM de 2.º grado de primaria a nivel de IE y estudiantes

2.º grado de primaria	Evaluados		Programados		Cobertura (%)	
	IE	alumnos	IE	alumnos	IE	alumnos
Nacional	5 987	172 940	5 976	165 658	99,8	95,8
Amazonas	238	3 932	238	3 837	100,0	97,6
Áncash	231	6 934	231	6 698	100,0	96,6
Apurímac	148	3 838	148	3 764	100,0	98,1
Arequipa	182	6 281	182	6 101	100,0	97,1
Ayacucho	184	5 114	184	5 016	100,0	98,1
Cajamarca	435	7 514	434	7 264	99,8	96,7
Callao	185	8 177	185	7 703	100,0	94,2
Cusco	254	8 696	253	8 544	99,6	98,3
Huancavelica	181	2 822	181	2 757	100,0	97,7
Huánuco	271	6 075	270	5 851	99,6	96,3
Ica	228	8 379	228	8 111	100,0	96,8
Junín	260	7 106	259	6 857	99,6	96,5
La Libertad	337	10 385	337	9 833	100,0	94,7
Lambayeque	317	11 024	317	10 455	100,0	94,8
Lima Metropolitana	244	9 770	244	9 293	100,0	95,1
Lima Provincia	206	6 754	206	6 470	100,0	95,8
Loreto	328	9 124	326	8 284	99,4	90,8
Madre de Dios	84	2 828	84	2 636	100,0	93,2
Moquegua	77	2 099	77	2 063	100,0	98,3
Pasco	201	3 652	199	3 470	99,0	95,0
Piura	421	12 869	420	12 403	99,8	96,4
Puno	244	6 895	244	6 813	100,0	98,8
San Martín	337	8 424	337	8 045	100,0	95,5
Tacna	90	3 511	90	3 481	100,0	99,1
Tumbes	116	3 762	116	3 584	100,0	95,3
Ucayali	188	6 975	186	6 325	98,9	90,7

Tabla 2.3 Cobertura de la EM de 4.º grado de primaria a nivel de IE y estudiantes

4.º grado de primaria	Evaluados		Programados		Cobertura (%)	
	IE	alumnos	IE	alumnos	IE	alumnos
Nacional	4 806	129 154	4 799	125 540	99,9	97,2
Amazonas	193	3 202	193	3 129	100,0	97,7
Áncash	221	6 798	221	6 619	100,0	97,4
Apurímac	152	3 933	152	3 878	100,0	98,6
Arequipa	92	3 145	92	3 091	100,0	98,3
Ayacucho	141	4 311	141	4 275	100,0	99,2
Cajamarca	561	8 734	560	8 491	99,8	97,2
Callao	99	4 403	99	4 288	100,0	97,4
Cusco	211	7 352	211	7 288	100,0	99,1
Huancavelica	217	3 168	217	3 123	100,0	98,6
Huánuco	248	5 521	248	5 357	100,0	97,0
Ica	125	5 125	125	5 017	100,0	97,9
Junín	254	6 231	251	6 071	98,8	97,4
La Libertad	236	6 280	236	6 097	100,0	97,1
Lambayeque	246	8 630	246	8 353	100,0	96,8
Lima Metropolitana	80	3 745	80	3 628	100,0	96,9
Lima Provincia	98	3 262	98	3 179	100,0	97,5
Loreto	295	7 476	295	6 965	100,0	93,2
Madre de Dios	51	1 948	50	1 880	98,0	96,5
Moquegua	68	2 050	68	2 034	100,0	99,2
Pasco	169	3 187	167	3 029	98,8	95,0
Piura	192	5 384	192	5 237	100,0	97,3
Puno	151	4 686	151	4 642	100,0	99,1
San Martín	430	10 267	430	9 952	100,0	96,9
Tacna	83	3 377	83	3 356	100,0	99,4
Tumbes	67	2 442	67	2 356	100,0	96,5
Ucayali	126	4 497	126	4 205	100,0	93,5

En relación con los niveles de inferencia de la EM 2019 en 2.º y 4.º grado de primaria, se pueden reportar resultados representativos en los siguientes niveles: a) nivel nacional; b) por sexo de los estudiantes; c) por gestión de las IE (estatales/no estatales); d) por ubicación geográfica de las IE (urbanas/rurales); e) por característica (polidocente completo/unidocente-multigrado); f) a nivel de DRE; y g) a nivel de gestión, área o característica dentro de la DRE.

Pesos muestrales EM 2019

Los pesos son ponderados que ayudan a corregir la distribución de la muestra en los estratos y a expandir la información muestral a la población.

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.

pIE_{ij} = 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.

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

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 y en la i-ésima IE.

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

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

Peso final. Se obtiene un peso para Lectura y otro para Matemática, cada cual dado por la siguiente fórmula:

$$pf_{kij} = pIE_{ij} \times psec_i \times a_{ki}$$

donde

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

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

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

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

Operativo de campo

Capítulo 3

La organización, la implementación y los procedimientos del operativo de campo de la ECE y EM se rigen por tres principios:

- mantener la confidencialidad de las pruebas;
- asegurar la estandarización de los procedimientos de la evaluación; y
- promover la probidad en las acciones de todos los actores durante todo el proceso de la evaluación.

Las estrategias propuestas para alcanzar estos principios se plasmaron en las especificaciones técnicas de los términos de referencia del servicio de aplicación de la ECE - EM. El Instituto Nacional de Estadística e Informática (INEI) es el operador logístico responsable de la ejecución de la aplicación de la ECE - EM en todo el país. Dicha institución tiene bajo su responsabilidad garantizar:

- la impresión, el modulado y el embalaje correctos de los instrumentos de evaluación, así como de los materiales de aplicación y capacitación;
- la distribución y la entrega oportunas y seguras de los materiales de capacitación e instrumentos de aplicación a cada una de las sedes a nivel nacional, así como el retorno de todos los instrumentos, aplicados y no aplicados, a la empresa responsable de la captura de datos;
- la seguridad y la confidencialidad de los instrumentos requeridos en todas las fases del operativo;
- la adecuada selección de la Red Administrativa (RA) según el perfil establecido;
- la capacitación de la RA en los procedimientos estandarizados de aplicación;
- el contacto previo con las IE para actualizar información de las IE y de los estudiantes; y
- la correcta aplicación de los instrumentos de evaluación en las IE.

Por su parte, el Minedu puso en práctica una estrategia de monitoreo y supervisión de todas las fases del operativo de campo. Para ello, contó con un equipo de 105 monitores Minedu, quienes supervisaron *in situ* una o más sedes provinciales. Dicho equipo tuvo como principales funciones asegurar que los procesos de capacitación cumplieran con el objetivo de transmitir de manera estandarizada los procedimientos de aplicación, velar por que el operador logístico llevara a cabo las acciones necesarias para asegurar la confidencialidad de los instrumentos de evaluación, y apoyar en la solución de los incidentes que se presentaran en la

sedes. Asimismo, este equipo fue el principal interlocutor con las instancias de gestión educativa descentralizada.

Además, por segundo año consecutivo se implementó para la ECE el modelo operativo que asegura que ninguna caja⁵ con instrumentos sea abierta antes del primer día de aplicación correspondiente a las fechas principales. Esto garantiza el cumplimiento de los principios de la evaluación (estandarización, confidencialidad y probidad).

El modelo consideró dos grandes momentos de aplicación:

- uno para IE programadas en **fechas principales** (es decir aquellas IE donde el aplicador⁶ retorna cada día de aplicación al local de la sede); y
- otro para IE programadas en **fechas posteriores**, una semana después de las fechas principales (estas IE son aquellas donde el aplicador debe pernoctar en una zona cercana a la IE, debido a la accesibilidad, tiempo de traslado, frecuencia de movilidad y la seguridad de la ruta a seguir).

En el caso de la EM, el modelo de aplicación considera una misma fecha de inicio de aplicación para todas las IE programadas, por lo que algunos aplicadores se desplazaron días antes al primer día de aplicación a fin de llegar en las fechas establecidas.

3.1 Organización territorial del operativo

Para gestionar territorialmente el operativo de campo, el país se dividió en 37 jurisdicciones regionales. La conformación de estas jurisdicciones regionales no se ciñe estrictamente a los límites del departamento donde se ubican; para conformarlas se tuvo en cuenta el criterio de accesibilidad. En cada jurisdicción regional de aplicación, se establecieron sedes provinciales y/o distritales, implementándose en total 274 locales de jurisdicción provinciales y distritales.

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

⁵Cada caja contiene los materiales necesarios para realizar la aplicación en una sección de una IE.

⁶Responsable de aplicar las pruebas a los estudiantes.

Tabla 3.1 Jurisdicciones regionales de aplicación y distribución de IE

Jurisdicción regional	ECE2S	EM2P	EM4P
AMAZONAS	788	419	407
ÁNCASH-CHIMBOTE	425	201	195
ÁNCASH-HUARAZ	577	129	132
APURÍMAC	471	212	214
AREQUIPA-NORTE	612	157	77
AREQUIPA-SUR	422	115	66
AYACUCHO	650	342	296
CAJAMARCA	1092	414	525
CALLAO	621	310	166
CUSCO	1045	372	313
HUANCAVELICA	476	244	266
HUANUCO	736	328	307
ICA	715	385	240
JUNÍN	1156	464	446
LA LIBERTAD- NOROESTE	661	211	136
LA LIBERTAD- SURESTE	761	271	259
LAMBAYEQUE	891	474	383
LIMA MET 1	971	80	82
LIMA MET 2	896	64	80
LIMA MET 3	677	49	51
LIMA MET 4	800	60	83
LIMA MET 5	757	68	72
LIMA MET 6	861	54	81
LIMA MET 7	628	41	38
LIMA-CAÑETE	267	87	47
LIMA-HUACHO	487	204	102
LORETO	680	331	315
MADRE DE DIOS	133	121	78
MOQUEGUA	126	112	103
PASCO	275	252	213
PIURA	1411	614	308
PUNO	1041	372	229
SAN MARTÍN-MOYOBAMBA	339	206	283
SAN MARTÍN-TARAPOTO	617	340	413
TACNA	241	145	140
TUMBES	182	159	100
UCAYALI	587	344	252
Total general	24 075	8751	7498

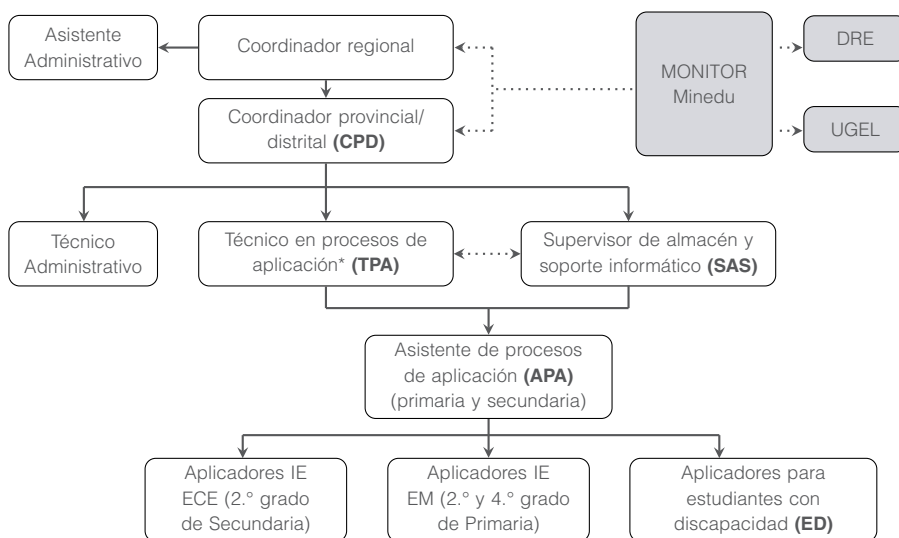
3.2 Conformación de la red administrativa de aplicación

El personal necesario para ejecutar el operativo de campo estuvo organizado en una Red Administrativa (RA) conformada por cuatro niveles de cargos.

Tabla 3.2 Personal del operativo

	Cargo	N.º de personas
Nivel I	Supervisores nacionales	50
	Coordinadores regionales	37
	Coordinadores provinciales distritales	274
Nivel II	Técnicos de procesos de aplicación	30
	Supervisores de almacén y soporte informático	274
Nivel III	Asistentes de procesos de aplicación ECE (2.º grado de secundaria) y EM (2.º y 4.º grado de primaria)	2 100
	Aplicadores de 2.º grado de primaria EM	10 668
Nivel IV	Aplicadores de 4.º grado de primaria EM	8 797
	Aplicadores de 2.º grado de secundaria ECE	42 846
	Aplicadores para atención de estudiantes con discapacidad	3 768

Figura 3.1 Organigrama de la Red Administrativa ECE - EM 2019



*Los técnicos de procesos de aplicación solo estarán presentes en las sedes provinciales/distritales que tengan más de 13 APA y serán el apoyo directo de los coordinadores provinciales/distritales.

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

- Los supervisores nacionales velan por el cumplimiento de los procesos y procedimientos del operativo en las diferentes regiones a las que son asignados. Además, garantizan las condiciones materiales necesarias para la evaluación y aseguran la consecución de personal en los diferentes niveles de la RA de las regiones a las que son asignados.
- Los coordinadores regionales, provinciales y distritales son responsables de las siguientes actividades: organizar, gestionar y dirigir el operativo, velar por el cumplimiento de los procedimientos y monitorear el desempeño de los miembros de la RA de su jurisdicción (región o provincia o distrito). Asimismo, son responsables de garantizar las condiciones materiales necesarias para el operativo y asegurar la implementación de medidas de seguridad en el almacén de instrumentos.
- Los técnicos de procesos de aplicación son responsables de supervisar e implementar los procesos de convocatoria, selección y capacitación de asistentes de procesos de aplicación. Además, trabajan directamente con los asistentes de procesos de aplicación, visitan las IE y monitorean el cumplimiento de los procedimientos los días de aplicación.
- Los supervisores de almacén y soporte informático son responsables de gestionar la recepción, el almacenamiento, el inventario y el despacho del material de aplicación, así como del uso de los sistemas de información implementados para estas actividades.
- Los asistentes de procesos de aplicación son responsables de ejecutar la selección y capacitación de los aplicadores, velar por el cumplimiento de los procedimientos, supervisar la aplicación y realizar el contacto previo con las IE asignadas.
- Los aplicadores son responsables de la adecuada aplicación de los instrumentos en las IE por evaluar.

Para cada uno de los niveles de la RA, el Minedu entregó manuales de aplicación y de funciones. En ese sentido, se elaboraron 16 manuales para el personal de la RA que atienden las particularidades de los grados evaluados.

- Manual del coordinador regional y provincial / distrital
- Manual del técnico de procesos de aplicación
- Manual del supervisor de almacén y soporte informático
- Manual del usuario del sistema
- Manual del técnico administrativo provincial
- Manual del asistente de procesos de aplicación

- Manual del aplicador 2.º grado de primaria EM
- Manual del aplicador 4.º grado de primaria EM
- Manual del aplicador 2.º grado de secundaria ECE
- Protocolo del aplicador de apoyo a estudiantes con discapacidad visual - baja visión
- Protocolo del aplicador de apoyo a estudiantes con discapacidad visual - ceguera
- Protocolo del aplicador de apoyo a estudiantes con discapacidad auditiva
- Protocolo del aplicador de apoyo a estudiantes con discapacidad motora
- 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 - prelector

3.3 Proceso de selección y capacitación del personal de la RA

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

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

El Minedu estableció, en los términos de referencia, el perfil profesional para cada uno de los cargos. Para los aplicadores, en la mayoría de los casos, los perfiles fueron los siguientes:

- Se consideró tres grupos:
 - I. Con experiencia en la ECE: (no necesita cumplir los otros requisitos del perfil)
Experiencia demostrada como aplicador en dos (2) de los operativos ECE 2015, 2016 o 2018, y que su desempeño no haya sido observado.
 - II. Sin experiencia en la ECE y con el siguiente perfil:
 - Docente con experiencia en aula, sin carga docente o administrativa en el sector estatal; o
 - Egresado universitario o egresado técnico, excepto carreras vinculadas al cuidado personal u oficios; o

- Estudiantes de institutos pedagógicos de los últimos tres años; o
- Estudiantes universitarios de carreras de humanidades de los últimos tres años.

III. Sin experiencia en la ECE y con práctica sustentada en:

- la aplicación de instrumentos estandarizados de evaluación educativa; o
- la aplicación de instrumentos de recojo de información; o
- participación en actividades con niñas, niños y adolescentes o en proyectos sociales.
- Con el siguiente perfil:
 - * Estudiantes universitarios de otras carreras que no sean de humanidades, de los últimos tres años; o
 - * Estudiantes de carreras técnicas del último año; o
 - * Estudiantes con estudios de CETPRO (excepto carreras vinculadas al cuidado personal u oficios)

Además, para los grupos I y II se considera otras características:

- Edad entre 19 y 60 años, salvo excepciones que deberán ser aprobadas por la UMC.
 - Deseable: habilidades para el buen manejo de grupos de niños y adolescentes.
 - Deseable: experiencia de trabajo con niños y niñas o adolescentes en aula, o animación sociocultural o aplicación de pruebas estandarizadas de rendimiento escolar.
 - Buen nivel de lectura comprensiva.
 - Adecuada dicción.
 - Obligatoriedad de asistir a las sesiones de capacitación.
 - Disposición a tiempo completo durante las fechas de capacitación y aplicación, y para viajar o trasladarse al lugar que se le designe.
 - Deseable: que disponga de teléfono móvil.
- Para el caso de aplicadores para la atención de estudiantes con discapacidad, se solicitó como mínimo (en orden de prioridad):
 - Egresado de carrera de educación especial; o
 - Estudiante universitario de carreras de educación especial de los dos últimos años; o

- Egresado de educación superior universitaria, titulado técnico con especialidad, diplomado en educación inclusiva o con experiencia en la atención de estudiantes con discapacidad; o
- Egresados técnicos o estudiantes del último año de la carrera de educación, enfermería, fisioterapia y rehabilitación o carreras afines; o
- Estudiantes universitarios de los dos últimos años o egresados de las carreras de psicología, educación, educación especial, trabajo social, sociología, antropología, comunicación social, enfermería u obstetricia o tecnología médica con mención en terapia física, terapia de lenguaje o terapia ocupacional o carreras afines; o
- Contar con:
 - * Una (1) experiencia de trabajo con niños y niñas, 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 autismo; deseable: manejo del sistema de comunicación por intercambio de tarjetas

Con el objetivo de desarrollar capacidades en los equipos de trabajo y de asegurar que recibieran las mismas instrucciones de aplicación en todo el país, se preparó un diseño de capacitación que fue replicado en todas las aulas de capacitación a nivel nacional. En dicha capacitación, la metodología incluyó actividades de lectura individual y grupal, ejercicios prácticos para el uso de las fichas de recojo de información y actividades sobre manejo de procedimientos y documentos de aplicación, así como actividades de simulación de las indicaciones para estudiantes en el aula de aplicación. También se detalló los indicadores de evaluación en cada una de las actividades propuestas. Al terminar la capacitación, los candidatos rindieron una prueba de salida orientada a conocer el nivel de comprensión de los procedimientos centrales de la aplicación, así como la habilidad de los candidatos para el manejo de los documentos e instrumentos de aplicación.

La capacitación de aplicadores se realizó en cada una de las sedes provinciales o distritales establecidas y estuvo a cargo de los asistentes de procesos de aplicación de la sede.

La capacitación tuvo una duración de dos días para los aplicadores de los operativos de 2.º y 4.º grado de primaria, así como para los de 2.º grado de secundaria. Las aulas estuvieron conformadas por un máximo de 30 candidatos. Además, los aplicadores que atendieron a estudiantes con discapacidad recibieron

medio día adicional de capacitación en sus propios manuales, según el tipo de discapacidad que iban atender.

Durante la capacitación, se instruyó a los candidatos en la aplicación estandarizada de instrumentos. Para ello, se utilizó el Manual del aplicador y los materiales elaborados para esta finalidad: fichas ópticas de práctica, papelógrafos, guiones de aplicación, entre otros.

3.4 Características de los instrumentos y documentos utilizados

Se utilizaron los siguientes instrumentos y documentos de aplicación:

2.º grado de primaria EM

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

4.º grado de primaria EM

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

2.º grado de secundaria ECE

- Cuadernillos integrados de pruebas ECE (un cuadernillo por día de prueba)
- Cuestionario para el estudiante
- Fichas de respuestas para estudiantes
- Registro de estudiantes
- Registro de préstamo
- Ficha de verificación
- Carta al director
- Cuestionario a la institución educativa
- Carta al docente
- Cuestionario para el docente de Comunicación
- Compromiso de confidencialidad para el docente (por día de aplicación)
- Ficha del aplicador de sección
- Ficha del aplicador líder
- Ficha del aplicador líder rural

Para facilitar y agilizar los procesos de aplicación, los cuadernillos de pruebas fueron personalizados con los nombres y apellidos de los estudiantes y un código de barras único. Para ello, se utilizó la información provista por el Siagie. El 95,1 % de los instrumentos enviados a las IE contó con información personalizada (por estudiante).

La inclusión de códigos y otros elementos de identificación en todos los instrumentos de esta evaluación permitió realizar la trazabilidad de cada uno de estos.

El operador logístico supervisó el avance y realizó el control de calidad de la impresión y el modulado de los instrumentos. El Minedu monitoreó estos procesos. Se embalaron 43 667 cajas con instrumentos (una caja para cada sección por evaluar y cajas con instrumentos adicionales por sede provincial y distrital). En todo momento, se garantizó la confidencialidad de los instrumentos.

Una vez moduladas, las cajas con los instrumentos fueron trasladadas en vehículos que aseguraron su arribo en condiciones adecuadas a las sedes provinciales y distritales, donde se hizo un inventario del número de cajas. Este traslado se realizó cumpliendo un cronograma y fue monitoreado por el personal del INEI y del Minedu. Cuando las cajas llegaron a cada sede provincial, fueron inventariadas y dispuestas en el almacén del local de jurisdicción provincial, el cual debía permanecer bajo llave y estar custodiado por la RA y con resguardo policial o privado hasta el final del operativo.

3.5 Proceso de aplicación

La duración y las fechas de aplicación de cada operativo de la ECE - EM 2019 se muestran en la tabla 3.3.

Tabla 3.3 Duración y fechas de la evaluación

Operativo	Duración	Fechas de evaluación	Criterio
2.º grado de primaria EM	2	27 y 28 de noviembre	<p>Un aplicador líder por IE con una sola sección programada que requiera pernocte de los aplicadores.</p> <p>Un aplicador por sección programada.</p> <p>Un aplicador para cada estudiante con discapacidad según registro de información UMC_ R-NEE.</p>
4.º grado de primaria EM	2	27 y 28 de noviembre	<p>Un aplicador líder por IE con una sola sección programada que requiera pernocte de los aplicadores.</p> <p>Un aplicador por sección programada.</p> <p>Un aplicador para cada estudiante con discapacidad según registro de información UMC_ R-NEE.</p>
2.º grado de secundaria ECE	3	5, 6 y 7 de noviembre*	<p>Mínimo un aplicador líder por IE con fechas principales de aplicación (según indicador de segmentación) que tengan dos o más secciones programadas.</p> <p>Un aplicador líder por IE con una sola sección programada, con fechas posteriores de aplicación (según indicador de segmentación).</p> <p>Un aplicador por sección programada.</p> <p>Un aplicador adicional para secciones de 30 a más estudiantes.</p> <p>Un aplicador para cada estudiante con discapacidad según registro de información UMC_ R-NEE.</p>

(* La ECE se aplicó en fechas principales (5, 6 y 7 de noviembre) en las IE sin pernocte y en fechas posteriores (12, 13 y 14 de noviembre) en las IE con pernocte.

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

2.º y 4.º grado de primaria EM

- I. Recepción de materiales. Los aplicadores de las IE que requieren pernocte se desplazan desde días antes al primer día de aplicación, para llegar a tiempo a las fechas programadas. En las IE que no requieren pernocte, el aplicador acude cada día de aplicación a la sede para recoger los instrumentos de evaluación.
- II. Presentación ante el director de la IE y organización de la aplicación. El aplicador llega temprano a la IE, se presenta ante el director, le entrega los documentos enviados por el Minedu y le solicita la nómina de matrícula del grado a evaluar

- y la lista auxiliar de asistencia del docente. El docente tutor firma el compromiso de confidencialidad.
- III. Preparación de los materiales antes de la evaluación. El aplicador solicita al director un espacio privado para preparar los cuadernillos y demás instrumentos. Solo y sin ayuda, el aplicador verifica la información de la nómina de matrícula con los nombres de los estudiantes de la FOAR y los cuadernillos.
 - IV. Preparación del aula e indicaciones. El aplicador organiza el mobiliario del aula, ubica a los estudiantes según el orden de la lista de estudiantes de la FOAR, entrega los cuadernillos de las pruebas y da lectura a las indicaciones de acuerdo con el guion de aplicación.
 - V. Orden y duración de las sesiones de aplicación. El aplicador sigue el guion de aplicación y administra los cuadernillos siguiendo la secuencia y los tiempos indicados en la tabla 3.4.

Tabla 3.4 Orden y duración de las sesiones de aplicación EM 2019

2.º grado de primaria EM		4.º grado de primaria EM	
Primer día	Segundo día	Primer día	Segundo día
Lectura (50 minutos)	Matemática (45 minutos)	Lectura (60 minutos)	Matemática (60 minutos)
Descanso (30 minutos)	Descanso (30 minutos)	Descanso (30 minutos)	Descanso (30 minutos)
Matemática (45 minutos)	Lectura (50 minutos)	Matemática (60 minutos)	Lectura (60 minutos)

- VI. Después de la aplicación, cada día, el aplicador se dirige a un ambiente privado dentro de la IE para realizar el traslado de las respuestas de los cuadernillos a la FOAR. Luego, el director verifica que el traslado haya sido correcto, respetando los procedimientos. Al culminar la revisión, el aplicador cuenta la totalidad de cuadernillos y procede a lacrarlos utilizando una bolsa de seguridad. El director presencia el conteo y firma un acta de conformidad, dando fe de la cantidad de cuadernillos que retornan al Minedu.

Al terminar las actividades previstas en la IE, el aplicador retorna inmediatamente a la sede para entregar todos los instrumentos de aplicación de la EM.

- VII. Al finalizar el último día de la aplicación, el director o personal responsable de la IE da conformidad a las condiciones de la aplicación en una declaración jurada de aplicación, la cual es firmada para cada sección evaluada. Para los casos en los que no se concretó la aplicación, se cuenta con una constancia de no aplicación, la cual detalla el motivo de la no aplicación, el mismo que es corroborado con una constancia de la UGEL.

Al culminar el trabajo en la IE, el aplicador se dirige a la sede donde el asistente de procesos verifica el material devuelto, teniendo en cuenta la cantidad y la

calidad de la información recogida. Luego de este proceso, los asistentes de procesos y los supervisores de almacén y soporte informático hacen un inventario de los instrumentos. Las fichas ópticas son retiradas de sus cajas, clasificadas y ordenadas según correlativo e instrumento. Los paquetes de cuadernillos retornan en sus mismas cajas.

2.º grado de secundaria ECE

- I. Recepción de materiales. En las IE que no requieren pernocte, el aplicador acude cada día de aplicación a la sede para recoger los instrumentos de evaluación. Los aplicadores de las IE que requieren pernocte se desplazan con el tiempo necesario para iniciar la aplicación en las fechas posteriores.
- II. Presentación ante el director de la IE y organización de la aplicación. El aplicador llega temprano a la IE, se presenta ante el director, le entrega los documentos enviados por el Minedu y le solicita la nómina de matrícula del grado a evaluar y la lista auxiliar de asistencia del docente. Se le solicita que designe a un docente para el apoyo en la organización de los estudiantes, y se le solicita que firme el compromiso de confidencialidad.
- III. Preparación del aula e indicaciones. El aplicador organiza el mobiliario del aula, ubica a los estudiantes según el orden del registro de estudiantes, entrega los cuadernillos de las pruebas y las fichas de respuestas, y luego da lectura a las indicaciones de acuerdo con el guion de aplicación.
- IV. Orden y duración de las sesiones de aplicación. El aplicador administra los cuadernillos siguiendo la secuencia y los tiempos indicados en la tabla 3.5.

Tabla 3.5 Orden y duración de las sesiones de aplicación 2.º grado de secundaria ECE

Grado y operativo	Primer día	Segundo día	Tercer día
2.º grado de secundaria ECE	Lectura (60 minutos)	Matemática (70 minutos)	Ciencia y Tecnología (90 minutos)
	Descanso (10 minutos)	Descanso (10 minutos)	Descanso (10 minutos)
	Matemática (70 minutos)	Lectura (60 minutos)	Cuestionario para el estudiante (30 minutos)

- V. Después de la aplicación, cada día, en un ambiente privado, el aplicador revisa las fichas de respuestas y organiza todos los instrumentos que deben ser devueltos al Minedu.
- VI. Al finalizar cada día de aplicación, el aplicador, en la dirección, cuenta la totalidad de cuadernillos y procede a lacrarlos utilizando una bolsa de

seguridad. El director, que presencia el conteo, firma un acta de conformidad respecto de las condiciones de aplicación, dando fe de la cantidad de cuadernillos que retornan al Minedu. Al terminar las actividades previstas en la IE, el aplicador retorna inmediatamente a la sede para entregar todos los instrumentos de aplicación de la ECE.

- VII. Al finalizar el último día de la aplicación, el director o personal responsable de la IE da conformidad a las condiciones de la aplicación en una declaración jurada de aplicación, la cual es firmada para cada sección evaluada. Para los casos en los que no se concretó la aplicación, se cuenta con una constancia de no aplicación, la cual detalla el motivo de la no aplicación, el mismo que es corroborado con una constancia de la UGEL.
- VIII. Al culminar el trabajo en la IE, el aplicador se dirige a la sede donde el asistente de procesos verifica el material devuelto, teniendo en cuenta la cantidad y la calidad de la información recogida. Luego de este proceso, los asistentes de procesos y los supervisores de almacén y soporte informático hacen un inventario de los instrumentos. Las fichas ópticas son retiradas de sus cajas, clasificadas y ordenadas según correlativo e instrumento. Los paquetes de cuadernillos retornan en sus mismas cajas.

Procesamiento de datos y análisis psicométrico

Capítulo 4

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). Ello permite garantizar la calidad de la información obtenida mediante la aplicación de sus evaluaciones. En este capítulo, se detalla cada uno de estos procesos.

4.1 Gestión y depuración de datos

Este proceso tiene como finalidad convertir la información contenida en los documentos físicos (fichas ópticas) en información digital (base de datos) consistente y confiable. Para ello, la UMC elabora especificaciones técnicas que detallan los procedimientos que se requieren para un correcto procesamiento de la información. Este proceso de depuración puede dividirse en tres etapas principales: gestión física, captura de datos e imágenes y depuración de bases de datos.

Luego, se contrata a una empresa que cuente con experiencia en este rubro y que siga los procedimientos bajo la atenta supervisión de la UMC. Dicha empresa provee todo el personal necesario, el cual es evaluado, seleccionado y capacitado por los representantes de la UMC.

4.1.1 Gestión física

Este proceso tiene como objetivo asegurar que toda la documentación física a procesar esté completa y lista para la captura de datos. En esta etapa, se realizan los siguientes pasos:

- **Recepción e inventario:** la empresa de captura de datos recibe todos los documentos físicos y realiza un inventario, escaneando el código de barras impreso en cada documento.
- **Clasificación:** se clasifican todos los documentos físicos según sus características, los cuales pueden ser fichas ópticas de respuesta, cuestionarios u otros.
- **Preparación:** consiste en agrupar los documentos físicos en lotes pequeños y quitar las grapas de estos para una mejor gestión en la siguiente etapa.
- **Almacenamiento:** todos los documentos físicos son almacenados por la empresa de captura de datos, de forma ordenada, para poder acceder a estos en caso se requiera durante la siguiente etapa.

Para asegurar una correcta gestión física de los documentos, la empresa de captura de datos proporciona un local con el espacio suficiente para todas las etapas, incluyendo un espacio exclusivo para el almacenamiento de todos los documentos debido a su confidencialidad.

4.1.2 Captura de datos

Este proceso tiene como objetivo digitalizar los documentos físicos y capturar los datos consignados en ellos.

- Digitalización y captura de datos: Consiste en pasar por un escáner especial todos los documentos físicos ya preparados. El escáner utilizado es capaz de digitalizar el documento físico en una imagen con gran resolución, así como capturar los datos consignados en este utilizando tres tecnologías: OMR (reconocimiento de marcas tipo burbujas), OCR (reconocimiento de caracteres impresos) e ICR (reconocimiento de caracteres escritos a mano).
- Control de calidad: Esta tarea se realiza cuando el software de reconocimiento de datos presenta un bajo nivel de confiabilidad de la data capturada. Estos casos son enviados a un grupo de personas dedicadas a revisar y asegurar que el dato capturado corresponda con el dato consignado en la imagen del documento.

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 exista coherencia entre los campos y la información adicional que se maneja en la evaluación. La UMC elabora un manual del depurador, el cual contiene todas las revisiones a realizar en cada campo de todas las bases de datos. Además de este manual, los depuradores son capacitados para que realicen de manera eficiente dicho proceso. Durante esta etapa de depuración, algunas de las revisiones realizadas son las siguientes:

- Revisiones iniciales:
 - Correspondencia de campos, según “diccionario de variables”, y la frecuencia de cada campo
 - 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 código modular y 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 el correlativo del estudiante
 - Duplicados en el correlativo del estudiante
 - Caracteres no válidos en el sexo del estudiante
 - Caracteres no válidos en la lengua materna del estudiante
 - Caracteres no válidos en la discapacidad del estudiante
 - Caracteres no válidos en el número de forma del cuadernillo
 - Caracteres no válidos en las respuestas del estudiante
 - Inconsistencia en las respuestas y la asistencia del estudiante
- Revisiones finales:
 - Repetir todas las revisiones iniciales
 - Contrastar con el listado de IE aplicadas, proporcionado por el equipo encargado del trabajo de campo de la UMC
 - Identificar valores fuera del rango permitido

4.2 Procesamiento psicométrico

4.2.1 Modelo Rasch para ítems dicotómicos

El análisis psicométrico de las pruebas aplicadas en el 2019 se basó en el modelo Rasch para ítems dicotómicos (Bond y Fox, 2015; Wright y Stone, 1998). Este modelo se centra en el análisis de cada ítem, concretamente en la interacción entre un ítem y una persona. 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 del ítem utilizado en términos de su dificultad (β)⁷. Por este motivo, se le denomina usualmente modelo de un parámetro (Hambleton, Swaminathan y Rogers, 1991). Es importante señalar que, en este reporte, se utilizará $\hat{\theta}$ para referirse a la cantidad del rasgo latente estimado para una persona y b para la dificultad estimada de un ítem en una muestra.

En los modelos Rasch, la habilidad de las personas y las dificultades de los ítems se ubican en la misma métrica. Al respecto, Smith y Kramer (1989) afirman que la existencia de una métrica común permite combinar la habilidad de la persona y la

⁷En 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.

dificultad del ítem para predecir el desempeño de esta persona en un ítem cualquiera e identificar respuestas inesperadas. El objetivo central del análisis Rasch es construir una escala conformada por los ítems ordenados según su dificultad. Ello implica que, a mayor habilidad, la persona presentará mayor probabilidad de acertar y, por lo tanto, obtendrá un mayor número de respuestas correctas. Es muy importante tener en cuenta que la medida estimada de la persona no es igual al puntaje directo que posee (número de ítems correctos); este puntaje es solo un insumo a partir del cual se construye la medida Rasch.

La relación entre la habilidad y la dificultad puede graficarse por medio de las curvas características del ítem (CCI), que brindan información concreta sobre la probabilidad de respuesta de una persona ante dicho ítem. Al trazar estas curvas, se establecen las siguientes relaciones en caso de tener ítems dicotómicos:

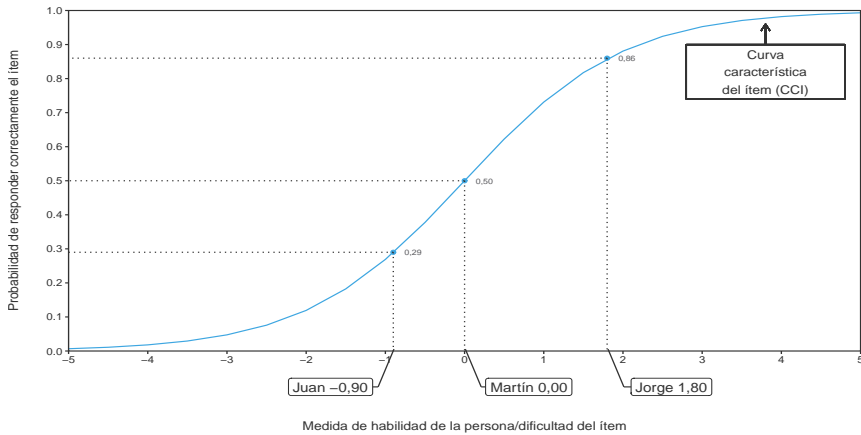
$$\begin{aligned}\theta > \beta; & \quad p(X = 1 \mid \theta, \beta) \in]0,5; 1,0] \\ \theta < \beta; & \quad p(X = 1 \mid \theta, \beta) \in [0,0; 0,5[\\ \theta = \beta; & \quad p(X = 1 \mid \theta, \beta) = 0,5\end{aligned}$$

En el primer caso, si la habilidad de la persona es mayor que la dificultad del ítem, la probabilidad de responder correctamente dicho 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 a 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:

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

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

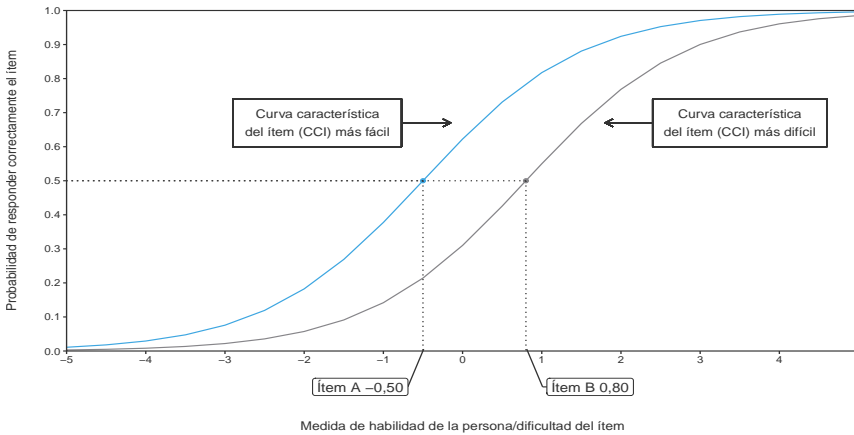
Figura 4.1 Curva característica del ítem y probabilidades de tres personas de responder correctamente dicho ítem



Se observa que, al enfrentarse a este ítem, Juan, cuya habilidad es de $-0,90$, tiene una probabilidad de $0,29$ de acertarlo; es decir, lo más probable es que lo 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 iguales posibilidades de acertar o 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, se expresa numéricamente la dificultad de un ítem como aquel 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 le pudiese evaluar en condiciones óptimas y con todos los ítems que potencialmente pueden usarse para medir dicho rasgo latente. Pero como ello 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 le 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, el cual se expresa mediante el error estándar de la dificultad del ítem (Wilson, 2005). Este error estándar puede ser calculado mediante esta fórmula:

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

4.3 Modelo Rasch de crédito parcial

Masters (1982) introdujo el modelo de crédito parcial para trabajar ítems politómicos de categorías ordenadas. Se deriva de los test 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 que puede aplicarse en test 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:

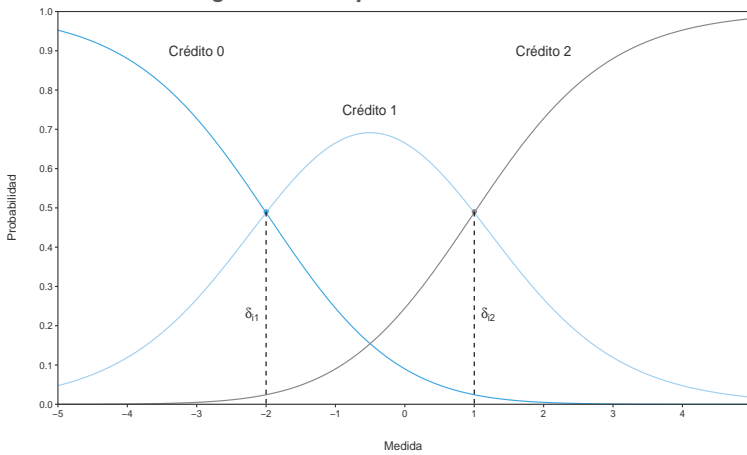
$$\sqrt{\frac{196}{4}} = x$$

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, todos los ítems y todas las 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 aquel valor en la escala del rasgo latente en el cual es igualmente probable recibir un puntaje o el inmediatamente adyacente. En algunos casos (Linacre, 2001), δ_{ij} es descompuesto en dos términos, D_i y F_{ij} , de tal manera que:

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

y

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

En este ejemplo, el primer umbral con un valor de $-1,74$ representa la medida en la cual es igualmente probable recibir 0 y 1 punto en el ítem. A partir de la figura 4.3, se puede señalar además que, si una persona tiene una medida del rasgo latente menor a $-1,74$, lo más probable es que haya recibido 0 puntos (crédito 0) en este ítem. Si su medida se encuentra entre $-1,71$ y $0,52$, lo más probable es que reciba 1 punto (crédito 1). Finalmente, 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:

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

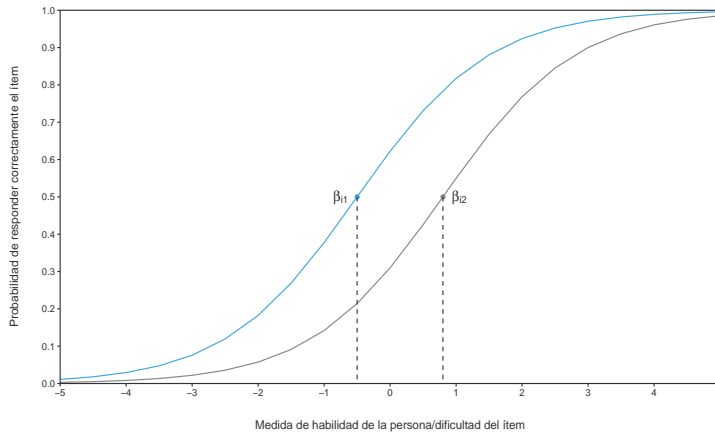
Supongamos que un ítem i tiene tres puntajes posibles, es decir, $X_{ni} = x \in (0,1,2)$. A partir de ello, se puede observar que $m_i = 2$. Por lo tanto, las probabilidades de obtener 0, 1 o 2 puntos se pueden expresar de la siguiente manera:

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

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

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

Como ya 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).

4.4 Análisis de ítems

El buen ajuste a un modelo es una parte importante en cualquier análisis estadístico, pero 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 con respecto al modelo suele dar lugar a la desestimación del modelo estadístico. Sin embargo, 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 Rasch. 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, en lugar de cómo el modelo se ajusta a los datos (Bond y Fox, 2015).

Como señala Wilson (2005), el ajuste de los ítems al modelo Rasch se fundamenta en dos estadísticos: el *outfit* e *infit*, cuyo cálculo se basa en los residuos cuadráticos estandarizados del modelo (Z_{ni}^2). A partir de las propuestas de diversos autores (Bond y Fox, 2015; Wilson, 2005; Wright y Linacre, 1994), se exponen brevemente el significado y el cálculo de ambos indicadores de ajuste.

El *outfit* (*outlier sensitive fit statistic*) es más sensible al comportamiento inesperado alejado de la medida de una persona. Se basa en la suma de cuadrados de los

residuales estandarizados (Z_{ni}^2), de tal manera que:

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

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

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

Como $R_{ni}^2 = VAR(P_{ni})Z_{ni}^2$ se puede redefinir el *infit* como una media cuadrática ponderada de residuales (Z_{ni}^2) que es sensible a patrones de respuesta irregulares:

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

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

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

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 hay un 25 % más de “ruido” en los datos 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, 2019):

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

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 (2019), 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, 2019).

Finalmente, se calcula la medida de dificultad de un ítem, utilizada por la teoría clásica de los test (*p*), que corresponde a la tasa de acierto. Sus valores varían entre 0 y 1, y los valores más bajos representan ítems más difíciles. Este índice para un ítem *i* se define como el cociente entre el número de personas que lo han acertado (A_i) y el número total de personas que lo han intentado resolver (N_i), tal como se muestra a continuación:

$$D_i = \frac{A_i}{N_i}$$

4.4.1 Calibración de ítems

En las tablas 4.1, 4.2 y 4.3 se muestra el detalle de las características psicométricas de los ítems de la prueba censal de 2.º grado de secundaria. Asimismo, en las tablas 4.4, 4.5, 4.6 y 4.7, se muestra las características psicométricas de los ítems para las pruebas muestrales de 2.º y 4.º grado de primaria.

Tabla 4.1 Dificultad y ajuste de los ítems al modelo Rasch de la prueba censal de Lectura de 2.º grado de secundaria

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B01	1	-1,864	0,007	0,92	0,79	0,27	0,95	1
B01	2	0,387	0,003	0,91	0,86	0,46	0,69	2
B01	3	-1,073	0,005	0,93	0,79	0,34	0,90	1
B01	4	-0,830	0,004	0,90	0,75	0,39	0,87	1
B01	5	0,746	0,003	1,02	1,04	0,37	0,61	2
B01	6	1,604	0,003	0,87	0,86	0,52	0,42	3
B01	7	2,624	0,033	0,81	0,65	0,56	0,23	4
B01	8	1,061	0,003	1,07	1,08	0,34	0,54	3
B01	9	0,435	0,003	0,99	0,99	0,39	0,68	2
B01	10	1,471	0,003	1,09	1,15	0,32	0,45	3
B01	11	-0,381	0,004	0,90	0,77	0,42	0,82	1
B01	12	1,610	0,003	1,06	1,09	0,35	0,42	3
B01	13	1,040	0,028	0,93	0,92	0,47	0,55	3
B01	14	2,071	0,003	0,97	0,98	0,42	0,33	4
B01	15_1	-0,354	0,003	1,08	1,07	0,36	0,70	1
B01	15_2	3,381	0,003	1,08	1,07	0,36	0,12	4
B01	16	1,308	0,003	0,99	0,99	0,42	0,49	3
B01	17	0,397	0,003	0,90	0,85	0,47	0,68	2
B01	18	0,769	0,003	0,85	0,78	0,54	0,61	2
B01	19	0,072	0,004	0,91	0,89	0,44	0,75	2
B02	1	-	-	-	-	-	-	-
B02	2	1,598	0,003	1,24	1,33	0,18	0,42	3

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B02	3	1,911	0,003	1,02	1,04	0,38	0,36	4
B02	4	1,131	0,003	0,94	0,93	0,46	0,53	3
B02	5	2,124	0,003	1,08	1,18	0,32	0,32	4
B02	6	0,858	0,003	1,06	1,07	0,34	0,59	2
B02	7	2,038	0,029	0,94	0,93	0,46	0,33	4
B02	8	0,928	0,003	1,15	1,21	0,26	0,57	2
B02	9	0,832	0,003	1,02	1,01	0,38	0,59	2
B02	10	0,322	0,003	1,00	0,93	0,38	0,70	2
B02	11	2,019	0,003	1,06	1,10	0,35	0,34	4
B02	12	3,038	0,004	1,08	1,22	0,27	0,17	4
B02	13	2,149	0,030	0,96	0,91	0,45	0,31	4
B02	14	1,138	0,003	1,02	1,05	0,38	0,53	3
B02	15	1,113	0,003	0,93	0,90	0,47	0,53	3
B02	16	1,523	0,003	1,08	1,11	0,34	0,44	3
B02	17	1,495	0,003	1,14	1,19	0,28	0,45	3
B02	18	1,650	0,003	0,96	0,97	0,44	0,41	3
B02	19	2,020	0,003	0,98	1,03	0,41	0,34	4
B03	1	0,531	0,007	0,88	0,81	0,50	0,66	2
B03	2	1,812	0,007	1,12	1,17	0,29	0,38	4
B03	3	0,509	0,007	0,95	0,91	0,44	0,66	2
B03	4	0,666	0,007	0,89	0,83	0,49	0,63	2
B03	5	1,763	0,007	1,16	1,21	0,26	0,39	3
B03	6	1,290	0,007	0,91	0,88	0,50	0,49	3
B04	1	2,238	0,008	0,85	0,80	0,53	0,29	4
B04	2	1,849	0,007	0,97	0,97	0,43	0,37	4
B04	3	1,476	0,007	1,05	1,05	0,36	0,45	3
B04	4_1	0,421	0,005	1,09	1,10	0,48	0,34	2
B04	4_2	1,676	0,005	1,09	1,10	0,48	0,38	3
B04	5_1	0,518	0,005	1,15	1,18	0,43	0,43	2
B04	5_2	2,209	0,005	1,15	1,18	0,43	0,26	4
B04	6	0,471	0,007	0,96	0,92	0,41	0,67	2
B05	1	1,978	0,007	1,14	1,20	0,27	0,34	4
B05	2	2,768	0,009	0,98	1,10	0,37	0,20	4
B05	3	1,898	0,007	1,00	1,03	0,40	0,36	4
B05	4	1,984	0,007	0,89	0,90	0,50	0,34	4
B05	5	2,180	0,008	0,97	0,98	0,42	0,30	4
B05	6	2,142	0,008	0,97	0,97	0,42	0,31	4
B06	1	0,064	0,008	0,96	1,01	0,38	0,75	2
B06	2	0,246	0,008	0,91	0,85	0,45	0,71	2
B06	3	1,753	0,007	0,91	0,90	0,48	0,39	3
B06	4	1,387	0,007	0,99	0,98	0,41	0,47	3
B06	5	0,417	0,007	0,85	0,78	0,52	0,68	2
B06	6	1,539	0,007	1,00	1,01	0,40	0,44	3
B07	1	1,199	0,007	1,11	1,14	0,32	0,51	3
B07	2	-0,344	0,009	0,86	0,69	0,47	0,81	1
B07	3	0,950	0,007	0,85	0,81	0,54	0,57	3

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B07	4	1,504	0,007	0,95	0,93	0,47	0,45	3
B07	5	1,908	0,007	0,94	0,98	0,46	0,36	4
B07	6	1,688	0,007	0,95	0,93	0,47	0,41	3
B08	1	0,572	0,007	0,87	0,82	0,51	0,65	2
B08	2	0,563	0,007	0,92	0,87	0,47	0,65	2
B08	3	0,480	0,007	1,08	1,12	0,31	0,67	2
B08	4	0,204	0,008	0,96	0,90	0,41	0,72	2
B08	5	1,063	0,007	0,98	0,96	0,43	0,54	3
B08	6	0,058	0,008	0,92	0,84	0,44	0,75	2
B09	1	0,446	0,007	0,90	0,85	0,47	0,67	2
B09	2	1,161	0,007	1,04	1,05	0,36	0,52	3
B09	3	0,799	0,007	0,94	0,91	0,45	0,60	2
B09	4	1,361	0,007	0,94	0,93	0,46	0,48	3
B09	5	0,476	0,007	1,07	1,11	0,30	0,67	2
B09	6	1,188	0,007	0,86	0,82	0,54	0,51	3
B10	1	1,536	0,007	1,03	1,04	0,38	0,44	3
B10	2	0,792	0,007	0,95	0,92	0,44	0,60	2
B10	3	1,692	0,007	0,96	0,96	0,44	0,40	3
B10	4	0,147	0,008	0,97	0,93	0,39	0,73	2
B10	5	1,756	0,007	0,93	0,94	0,47	0,39	3
B10	6	-	-	-	-	-	-	-
B11	1	0,528	0,007	0,95	0,91	0,43	0,66	2
B11	2	1,669	0,007	1,25	1,34	0,16	0,41	3
B11	3	0,959	0,007	1,06	1,10	0,33	0,57	3
B11	4	-	-	-	-	-	-	-
B11	5	1,475	0,007	0,99	0,99	0,41	0,45	3
B11	6	1,314	0,007	1,09	1,12	0,32	0,49	3
B12	1	-0,340	0,009	0,91	0,83	0,42	0,81	1
B12	2	1,767	0,007	0,90	0,89	0,51	0,39	4
B12	3	-0,089	0,008	0,85	0,71	0,50	0,77	1
B12	4	-0,540	0,009	0,88	0,74	0,43	0,84	1
B12	5	0,772	0,007	1,00	0,99	0,41	0,61	2
B12	6	1,183	0,007	0,92	0,91	0,49	0,52	3

Tabla 4.2 Dificultad y ajuste de los ítems al modelo Rasch de la prueba censal de Matemática de 2.º grado de secundaria

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B01	1	-	-	-	-	-	-	-
B01	2	1,327	0,007	1,15	1,30	0,38	0,51	3
B01	3	1,851	0,007	1,06	1,08	0,46	0,41	3
B01	4	-	-	-	-	-	-	-
B01	5	1,117	0,007	1,02	1,01	0,47	0,56	2
B02	1	-	-	-	-	-	-	-

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B02	2	1,426	0,007	1,12	1,19	0,41	0,49	3
B02	3	2,337	0,008	0,99	1,09	0,50	0,32	4
B02	4	-	-	-	-	-	-	-
B02	5	0,731	0,007	0,96	0,95	0,48	0,62	2
B03	1	0,422	0,008	1,01	1,02	0,44	0,69	2
B03	2	0,460	0,008	0,89	0,87	0,51	0,68	2
B03	3	-	-	-	-	-	-	-
B03	4	2,376	0,008	1,13	1,33	0,41	0,32	4
B03	5	-0,104	0,008	0,90	0,84	0,46	0,77	1
B04	1	0,326	0,008	1,00	0,91	0,45	0,69	2
B04	2	0,160	0,008	0,92	0,94	0,47	0,72	1
B04	3	0,606	0,008	0,87	0,84	0,54	0,64	2
B04	4	-	-	-	-	-	-	-
B04	5	1,709	0,007	1,02	1,07	0,49	0,43	3
B05	1	-0,225	0,009	0,82	0,67	0,51	0,79	1
B05	2	1,763	0,007	1,12	1,20	0,41	0,43	3
B05	3	-	-	-	-	-	-	-
B05	4	1,298	0,007	1,03	1,06	0,45	0,52	2
B05	5	1,098	0,007	0,96	0,96	0,50	0,56	2
B06	1	-0,534	0,009	0,89	0,81	0,43	0,83	1
B06	2	-	-	-	-	-	-	-
B06	3	2,581	0,008	0,99	1,19	0,48	0,28	4
B06	4	0,764	0,007	1,06	1,06	0,42	0,62	2
B06	5	1,020	0,007	0,90	0,87	0,54	0,57	2
B07	1	0,416	0,008	0,92	0,92	0,49	0,68	2
B07	2	-1,008	0,01	0,93	1,11	0,36	0,88	1
B07	3	2,069	0,008	0,97	1,07	0,51	0,36	4
B07	4	-	-	-	-	-	-	-
B07	5	-	-	-	-	-	-	-
B08	1	0,387	0,008	0,97	0,93	0,47	0,69	2
B08	2	1,092	0,007	0,87	0,82	0,57	0,56	2
B08	3	1,449	0,007	1,08	1,11	0,45	0,49	3
B08	4	1,399	0,007	0,88	0,84	0,57	0,50	3
B08	5	1,900	0,008	1,01	1,07	0,49	0,40	3
B09	1	0,987	0,007	1,00	1,00	0,48	0,57	2
B09	2	2,019	0,008	1,11	1,25	0,44	0,38	3
B09	3	2,268	0,008	0,94	0,99	0,54	0,33	4
B09	4	-	-	-	-	-	-	-
B09	5	0,746	0,008	0,92	0,88	0,52	0,62	2
B10	1	1,123	0,007	1,14	1,24	0,37	0,55	2
B10	2	2,500	0,008	1,04	1,29	0,44	0,29	4
B10	3	-	-	-	-	-	-	-
B10	4	1,572	0,007	1,04	1,07	0,46	0,46	3
B10	5	1,265	0,007	0,96	0,97	0,50	0,52	2
B11	1	-0,733	0,004	1,02	1,10	0,33	0,85	1
B11	2	-0,291	0,004	0,83	0,71	0,50	0,80	1

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B12	1	0,859	0,003	1,14	1,20	0,38	0,60	2
B12	2	0,269	0,004	0,93	0,88	0,48	0,71	2
B12	3	1,148	0,003	1,08	1,12	0,43	0,55	2
B12	4	2,597	0,004	0,86	0,90	0,57	0,28	4
B12	5	-	-	-	-	-	-	-
B12	6	-0,869	0,005	1,00	1,30	0,31	0,87	1
B12	7	1,469	0,003	1,15	1,23	0,39	0,48	3
B13	1	-0,369	0,004	1,01	1,06	0,37	0,81	1
B13	2	0,544	0,003	0,91	0,88	0,50	0,66	2
B14	1	-1,253	0,005	0,93	1,04	0,33	0,90	1
B14	2	-	-	-	-	-	-	-
B14	3	-	-	-	-	-	-	-
B14	4	-	-	-	-	-	-	-
B14	5	-0,563	0,004	1,00	1,28	0,34	0,83	1
B14	6	-	-	-	-	-	-	-
B14	7	0,353	0,003	0,98	0,98	0,45	0,7	2
B15	1	-0,200	0,004	0,87	0,74	0,48	0,78	1
B15	2	-0,332	0,004	1,05	1,27	0,34	0,80	1
B16	1	-0,743	0,004	0,90	0,93	0,40	0,85	1
B16	2	0,365	0,003	1,01	0,97	0,44	0,69	2
B16	3	0,709	0,003	1,03	1,01	0,44	0,63	2
B16	4	1,901	0,003	0,87	0,90	0,57	0,40	3
B16	5	0,341	0,003	1,07	1,10	0,39	0,70	2
B16	6	0,980	0,003	1,14	1,26	0,38	0,58	2
B16	7	-0,859	0,004	1,00	0,95	0,34	0,86	1
B17	1	-0,445	0,004	0,96	1,00	0,39	0,82	1
B17	2	0,784	0,003	0,97	0,94	0,49	0,61	2
B18	1	-1,336	0,005	0,97	1,29	0,29	0,91	1
B18	2	-	-	-	-	-	-	-
B18	3	-0,347	0,004	0,96	1,04	0,40	0,80	1
B18	4	2,160	0,003	0,96	1,06	0,52	0,35	4
B18	5	2,195	0,003	0,98	1,03	0,51	0,35	4
B18	6	2,107	0,003	1,00	1,11	0,49	0,36	4
B18	7	0,471	0,003	1,12	1,23	0,36	0,67	2
B19	1	3,894	0,044	0,93	0,72	0,48	0,12	4
B20	1	-	-	-	-	-	-	-
B21	1	3,831	0,043	0,94	0,73	0,47	0,13	4
B22	1	3,277	0,038	0,92	0,80	0,52	0,19	4

Tabla 4.3 Dificultad y ajuste de los ítems al modelo Rasch de la prueba censal de Ciencia y Tecnología de 2.º grado de secundaria

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B01	1	-0,199	0,005	0,87	0,85	0,49	0,55	3
B01	2	-1,115	0,006	0,90	0,84	0,44	0,74	3
B01	3	-0,672	0,005	0,87	0,82	0,49	0,65	4
B01	4	1,412	0,006	1,03	1,15	0,19	0,22	2
B01	5	-0,332	0,005	1,15	1,18	0,11	0,58	3
B01	6	1,865	0,007	1,08	1,26	0,09	0,16	4
B01	7	0,930	0,006	1,05	1,10	0,21	0,31	3
(B01,B07)	(8_1,1_1)	0,981	0,024	0,83	0,67	0,53	0,09	3
(B01,B07)	(8_2,1_2)	1,344	0,024	0,83	0,67	0,53	0,14	3
B01	9	1,386	0,006	1,06	1,19	0,16	0,23	4
B01	10	1,227	0,006	1,12	1,25	0,09	0,25	3
B01	11	-0,333	0,005	0,94	0,93	0,40	0,58	3
B01	12	-0,224	0,005	0,92	0,91	0,42	0,56	3
B01	13	1,058	0,006	1,01	1,05	0,26	0,29	3
B01	14	0,382	0,005	0,98	0,98	0,33	0,42	4
B01	15	0,893	0,006	1,03	1,05	0,25	0,32	4
B01	16	-0,303	0,005	0,98	0,97	0,34	0,57	3
B02	1	1,150	0,006	1,15	1,31	0,02	0,26	4
B02	2	0,162	0,005	1,01	1,02	0,28	0,46	2
B02	3	-0,991	0,006	0,97	0,95	0,33	0,71	3
B02	4	-2,565	0,009	0,94	0,80	0,31	0,91	4
B02	5	0,219	0,005	0,92	0,91	0,42	0,45	3
B02	6	-1,932	0,007	0,92	0,79	0,38	0,85	1
B02	7	-1,557	0,006	0,92	0,84	0,39	0,80	2
B02	8	-0,335	0,005	1,05	1,06	0,23	0,57	4
B02	9	1,427	0,006	1,09	1,25	0,08	0,21	4
B02	10	0,312	0,005	1,11	1,13	0,14	0,43	3
B02	11	0,306	0,005	1,11	1,13	0,14	0,43	4
B02	12	-0,157	0,005	0,88	0,86	0,48	0,53	3
B02	13	0,137	0,005	1,09	1,11	0,17	0,47	3
B02	14	0,028	0,005	1,11	1,14	0,14	0,49	3
B02	15	-0,771	0,005	0,92	0,89	0,41	0,66	3
B03	1	0,300	0,005	1,05	1,06	0,25	0,43	4
B03	2	0,848	0,006	1,00	1,02	0,29	0,32	4
B03	3	-0,303	0,005	0,99	0,99	0,33	0,57	3
B03	4	0,132	0,005	1,05	1,06	0,25	0,47	3
B03	5	0,045	0,005	0,92	0,91	0,42	0,49	3
B03	6	-0,186	0,005	1,00	1,00	0,32	0,54	2
B03	7	0,470	0,005	0,98	0,98	0,34	0,40	2
B03	8	-0,353	0,005	0,96	0,95	0,37	0,58	3
B03	9	0,488	0,005	0,92	0,91	0,42	0,39	3
B03	10	0,707	0,005	0,95	0,95	0,37	0,35	4
B03	11	0,012	0,005	1,02	1,02	0,30	0,50	3

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B03	12	-2,372	0,008	0,92	0,76	0,35	0,89	3
B03	13	0,414	0,005	1,06	1,08	0,23	0,41	2
B03	14	-1,207	0,006	0,93	0,88	0,39	0,75	4
B03	15	-0,256	0,005	1,03	1,05	0,27	0,56	4
B04	1	-1,018	0,006	0,89	0,84	0,44	0,72	4
B04	2	0,172	0,005	1,01	1,01	0,32	0,47	3
B04	3	0,039	0,005	1,05	1,06	0,27	0,50	4
B04	4	1,593	0,006	1,07	1,18	0,17	0,20	4
B04	5	0,747	0,005	1,10	1,14	0,18	0,35	1
B04	6	0,674	0,005	1,05	1,08	0,25	0,36	3
B04	7	0,100	0,005	0,99	0,99	0,34	0,49	4
B04	8	0,976	0,050	0,92	0,88	0,42	0,31	2
B04	9	0,736	0,005	1,00	1,03	0,30	0,35	1
B04	10	-0,689	0,005	0,93	0,90	0,41	0,66	3
B04	11	-0,075	0,005	0,98	0,97	0,36	0,53	1
B04	12	-2,378	0,008	0,93	0,75	0,33	0,90	1
B04	13	-0,432	0,005	1,01	1,04	0,30	0,60	3
B04	14	0,679	0,005	1,02	1,05	0,28	0,36	3
B04	15	1,467	0,006	1,04	1,11	0,22	0,22	2
B04	16	0,328	0,005	0,97	0,97	0,36	0,44	3
B05	1	-1,294	0,006	0,93	0,86	0,39	0,77	3
B05	2	-0,238	0,005	1,00	0,99	0,34	0,57	3
B05	3	-0,256	0,005	0,96	0,95	0,39	0,58	2
B05	4	-0,863	0,006	0,87	0,80	0,48	0,70	1
B05	5	0,363	0,005	1,00	1,01	0,33	0,44	1
B05	6	0,401	0,005	1,13	1,16	0,17	0,43	1
B05	7	-0,042	0,005	1,14	1,17	0,16	0,53	3
B05	8_1	1,553	0,052	0,92	0,80	0,40	0,18	3
B05	8_2	3,361	0,052	0,92	0,80	0,40	0,03	3
B05	9	-0,819	0,006	0,91	0,86	0,43	0,69	4
B05	10	0,973	0,006	1,09	1,18	0,19	0,32	4
B05	11	1,345	0,006	1,10	1,21	0,16	0,25	4
B05	12	-0,002	0,005	1,03	1,04	0,30	0,52	1
B05	13	-1,135	0,006	0,96	0,95	0,35	0,75	3
B05	14	1,186	0,006	1,05	1,13	0,23	0,28	2
B05	15	-1,178	0,006	0,86	0,76	0,49	0,75	2
B05	16	-0,332	0,005	0,94	0,92	0,40	0,59	3
B06	1	-0,675	0,006	0,88	0,84	0,47	0,66	4
B06	2	0,427	0,005	1,04	1,05	0,28	0,43	2
B06	3	0,303	0,005	1,07	1,09	0,25	0,45	4
B06	4	-2,024	0,008	0,98	0,95	0,28	0,87	3
B06	5	0,388	0,005	1,06	1,09	0,25	0,43	3
B06	6	-	-	-	-	-	-	-
B06	7	0,796	0,005	1,12	1,18	0,17	0,35	4
(B06,B08)	(8,1)	1,671	0,041	0,92	0,82	0,37	0,19	3
B06	9	-0,193	0,005	0,93	0,92	0,42	0,56	4

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B06	10	-1,576	0,007	0,95	0,87	0,36	0,81	3
B06	11	-1,452	0,006	0,99	0,99	0,30	0,79	2
B06	12	-0,111	0,005	0,97	0,96	0,37	0,54	3
B06	13	1,038	0,006	0,93	0,93	0,39	0,30	4
B06	14	-0,079	0,005	0,93	0,91	0,43	0,54	4
B06	15	-2,143	0,008	0,93	0,78	0,36	0,88	2
B06	16	-1,720	0,007	0,98	1,02	0,29	0,83	1

Tabla 4.4 Dificultad y ajuste de los ítems al modelo Rasch de la prueba muestral de Lectura de 2.º grado de primaria

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B01	1	-0,674	0,010	0,90	0,68	0,35	0,93	1
B01	2	-0,565	0,010	0,87	0,56	0,39	0,92	1
B01	3	-0,274	0,009	0,86	0,64	0,41	0,90	1
B01	4	1,115	0,006	0,97	0,96	0,41	0,72	2
B01	5	2,425	0,006	1,16	1,24	0,29	0,46	3
B01	6	1,780	0,006	0,96	0,94	0,45	0,59	2
B01	7	2,518	0,006	1,20	1,29	0,26	0,44	3
B01	8	1,478	0,006	0,95	0,89	0,46	0,65	2
B01	9	1,099	0,006	1,07	1,17	0,32	0,72	2
B01	10	-	-	-	-	-	-	-
B01	11	2,125	0,005	1,00	0,99	0,43	0,52	3
B01	12	1,377	0,006	0,92	0,87	0,48	0,67	2
B01	13	1,450	0,006	0,89	0,83	0,50	0,66	2
B01	14	1,353	0,006	0,97	0,90	0,44	0,67	2
B02	1	0,298	0,007	0,89	0,71	0,44	0,84	2
B02	2	0,391	0,007	0,87	0,71	0,46	0,83	2
B02	3	2,015	0,005	1,04	1,05	0,39	0,54	2
B02	4	1,730	0,006	0,92	0,92	0,48	0,60	2
B02	5	1,183	0,006	0,92	0,84	0,47	0,70	2
B02	6	0,876	0,006	0,83	0,71	0,52	0,76	2
B02	7	0,866	0,006	0,84	0,70	0,52	0,76	2
B02	8	2,342	0,005	1,13	1,18	0,32	0,47	3
B02	9	1,965	0,005	1,14	1,23	0,30	0,55	2
B02	10	1,045	0,006	0,92	0,85	0,46	0,73	2
(B03,B08)	(1,1)	1,878	0,008	0,89	0,85	0,52	0,57	2
(B03,B08)	(2,3)	2,601	0,008	1,03	1,07	0,41	0,42	3
(B03,B08)	(3,4)	2,597	0,008	0,96	0,99	0,46	0,42	3
B03	4	3,211	0,009	1,06	1,12	0,35	0,34	3
B03	5	2,799	0,009	1,16	1,23	0,28	0,42	3
B03	6	2,608	0,009	1,15	1,22	0,29	0,46	3
B04	1	0,906	0,009	0,85	0,73	0,51	0,75	2
B04	2	1,947	0,008	1,06	1,06	0,39	0,56	2

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B04	3	2,384	0,008	1,19	1,25	0,29	0,47	3
B04	4	-	-	-	-	-	-	-
B04	5	2,537	0,008	1,10	1,16	0,35	0,44	3
(B05,B11)	(1,1)	2,865	0,008	1,16	1,28	0,26	0,36	3
(B05,B11)	(2,2)	1,884	0,008	0,99	0,98	0,42	0,57	2
(B05,B11)	(3,3)	2,459	0,008	1,07	1,10	0,35	0,44	3
(B05,B11)	(4,4)	2,190	0,008	0,94	0,92	0,47	0,50	3
(B05,B11)	(5,5)	2,262	0,008	1,14	1,17	0,30	0,49	3
B05	6	2,225	0,009	1,03	1,05	0,37	0,53	3
B06	1	1,793	0,009	0,97	0,94	0,41	0,62	2
B06	2	2,818	0,009	1,06	1,11	0,34	0,40	3
B06	3	2,851	0,009	1,07	1,10	0,33	0,39	3
B06	4	2,645	0,009	1,14	1,19	0,27	0,44	3
B06	5	3,553	0,010	1,10	1,20	0,28	0,26	3
B07	1	-0,091	0,021	0,98	0,95	0,34	0,81	1
B08	2	2,028	0,017	0,99	0,99	0,42	0,40	3
B08	5	2,102	0,017	1,15	1,17	0,29	0,39	3
B09	1	0,858	0,017	0,90	0,84	0,48	0,65	2
B09	2	2,060	0,017	0,94	0,93	0,46	0,40	3
B09	3	1,752	0,017	0,93	0,92	0,47	0,46	2
B09	4	1,862	0,017	0,94	0,93	0,47	0,44	2
B09	5	1,675	0,017	1,04	1,06	0,38	0,48	2
B10	1	0,334	0,019	0,92	0,88	0,43	0,75	2
B12	1	1,939	0,006	0,98	0,96	0,43	0,59	2
B12	2	2,693	0,006	1,16	1,21	0,27	0,43	3
B12	3	2,158	0,006	0,94	0,93	0,46	0,55	3
B12	4	1,381	0,007	0,83	0,75	0,53	0,70	2
B13	1	0,770	0,011	0,84	0,69	0,49	0,80	2
B13	2	1,302	0,010	0,92	0,83	0,46	0,72	2
B13	3	1,035	0,010	0,88	0,76	0,48	0,76	2
B13	4	1,131	0,010	0,91	0,79	0,46	0,75	2
B13	5	2,063	0,009	0,97	0,95	0,45	0,57	3
B13	6	2,469	0,009	1,09	1,14	0,35	0,48	3
B14	1	3,020	0,009	1,16	1,30	0,27	0,37	3
(B14,B22)	(2,1)	2,430	0,008	1,06	1,12	0,39	0,46	3
(B14,B22)	(3,3)	2,530	0,008	1,08	1,12	0,37	0,44	3
(B14,B22)	(4,4)	2,773	0,008	0,95	1,00	0,47	0,39	3
B14	5	2,379	0,009	1,03	1,03	0,40	0,50	3
(B15,B19)	(1,1)	2,565	0,008	1,07	1,16	0,35	0,42	3
(B15,B19)	(2,2)	2,593	0,008	1,02	1,08	0,39	0,42	3
(B15,B19)	(3,4)	1,676	0,008	0,97	0,95	0,44	0,61	2
B15	4	3,212	0,009	1,09	1,21	0,29	0,32	3
B15	5	2,960	0,009	1,04	1,09	0,35	0,37	3
(B15,B19)	(6,5)	3,155	0,008	1,14	1,24	0,28	0,31	3
B16	1	1,703	0,008	0,98	0,96	0,43	0,60	2
B16	2	2,723	0,008	1,10	1,19	0,32	0,39	3

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B16	3	2,365	0,008	0,97	0,97	0,45	0,46	3
B16	4	1,839	0,008	1,07	1,09	0,35	0,57	2
B16	5	1,723	0,008	0,96	0,97	0,44	0,60	2
B17	1	0,564	0,018	0,92	0,88	0,44	0,70	2
B18	1	0,843	0,012	0,89	0,82	0,49	0,65	2
B18	2	0,840	0,012	0,86	0,79	0,51	0,65	2
B18	3	0,288	0,013	0,84	0,70	0,50	0,75	2
B18	4	1,638	0,012	1,13	1,19	0,29	0,48	2
B19	3	1,192	0,017	0,97	0,92	0,43	0,58	2
B20	1	0,552	0,018	0,89	0,82	0,48	0,71	2
B21	1	0,880	0,017	1,02	0,98	0,38	0,64	2
B21	2	1,589	0,017	0,98	0,96	0,43	0,49	2
B21	3	1,581	0,017	1,08	1,09	0,34	0,50	2
B21	4	1,242	0,017	0,89	0,84	0,51	0,57	2
B21	5	1,224	0,017	0,91	0,87	0,49	0,57	2
B22	2	2,090	0,017	1,07	1,10	0,35	0,39	3
B22	5	2,174	0,017	1,23	1,33	0,20	0,37	3

Tabla 4.5 Dificultad y ajuste de los ítems al modelo Rasch de la prueba muestral de Matemática de 2.º grado de primaria

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B01	1	-	-	-	-	-	-	-
B01	2	2,958	0,010	1,10	1,25	0,37	0,25	3
B01	3	1,161	0,009	0,87	0,84	0,55	0,47	2
B01	4	2,325	0,010	1,05	1,10	0,42	0,32	3
B01	5	2,931	0,010	1,25	1,42	0,26	0,25	3
B01	6	1,624	0,009	0,82	0,78	0,60	0,41	2
B01	7	2,411	0,010	0,99	1,02	0,46	0,31	3
B02	1	1,704	0,009	0,89	0,87	0,54	0,40	2
B02	2	3,527	0,011	1,04	1,16	0,39	0,20	3
B02	3	2,463	0,010	1,22	1,38	0,29	0,31	3
B02	4	1,895	0,009	1,07	1,11	0,41	0,38	3
B02	5	2,246	0,009	1,09	1,15	0,39	0,33	3
B02	6	2,719	0,010	1,09	1,12	0,39	0,28	3
B02	7	2,264	0,009	1,04	1,09	0,43	0,33	3
B03	1	-1,987	0,020	0,95	0,91	0,38	0,78	1
B03	2	-0,379	0,017	0,82	0,77	0,56	0,57	1
B03	3	-1,229	0,018	0,91	0,88	0,45	0,69	1
B03	4	0,411	0,017	0,93	0,91	0,49	0,46	2
B03	5	0,390	0,017	0,83	0,78	0,57	0,46	2
B03	6	-1,041	0,018	0,94	0,91	0,44	0,66	1
B03	7	-0,593	0,017	0,98	1,01	0,42	0,60	1
B04	1	-1,489	0,018	0,89	0,83	0,46	0,72	1

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B04	2	-1,553	0,019	0,93	0,86	0,43	0,72	1
B04	3	-0,104	0,017	0,96	0,95	0,46	0,53	1
B04	4	0,126	0,017	0,85	0,80	0,56	0,49	2
B04	5	1,134	0,018	1,01	1,03	0,44	0,36	2
B04	6	-1,288	0,018	0,92	0,87	0,45	0,69	1
B04	7	-1,133	0,018	0,99	0,93	0,41	0,67	1
B05	1	2,247	0,010	1,07	1,12	0,41	0,33	3
B05	2	1,841	0,009	0,93	0,90	0,52	0,38	2
B05	3	1,739	0,009	1,10	1,12	0,40	0,39	2
B05	4	1,941	0,009	1,04	1,04	0,44	0,36	3
B05	5	1,880	0,009	1,14	1,19	0,36	0,37	3
B05	6	2,175	0,009	0,98	0,96	0,49	0,34	3
B05	7	2,248	0,010	1,06	1,13	0,42	0,33	3
B06	1	1,979	0,009	1,07	1,14	0,41	0,37	3
B06	2	2,120	0,009	0,85	0,79	0,58	0,35	3
B06	3	2,273	0,010	1,07	1,09	0,42	0,33	3
B06	4	2,938	0,010	1,05	1,02	0,43	0,25	3
B06	5	3,517	0,011	1,04	1,11	0,40	0,20	3
B06	6	2,648	0,010	1,03	1,07	0,43	0,28	3
B06	7	1,725	0,009	0,77	0,72	0,64	0,40	2
B07	1	-3,328	0,026	0,97	0,95	0,30	0,89	1
B07	2	-1,429	0,018	0,99	1,03	0,38	0,71	1
B07	3	-0,699	0,017	0,78	0,71	0,58	0,61	1
B07	4	0,652	0,017	0,86	0,83	0,55	0,42	2
B07	5	-0,054	0,017	0,90	0,88	0,50	0,52	1
B07	6	0,076	0,017	0,91	0,92	0,49	0,50	1
B07	7	-1,452	0,018	1,13	1,17	0,27	0,71	1
B08	1	-1,781	0,019	0,91	0,88	0,43	0,75	1
B08	2	-1,836	0,019	0,95	0,92	0,40	0,75	1
B08	3	-0,297	0,017	0,97	0,96	0,45	0,55	1
B08	4	0,543	0,017	0,75	0,69	0,64	0,43	2
B08	5	0,610	0,017	0,79	0,73	0,61	0,42	2
B08	6	0,252	0,017	0,80	0,75	0,60	0,47	2
B08	7	-1,065	0,018	0,95	0,96	0,44	0,66	1
B09	1	-0,840	0,006	0,96	0,94	0,42	0,72	1
B09	2	-0,294	0,006	1,01	1,11	0,39	0,65	1
B09	3	0,569	0,006	1,04	1,02	0,42	0,53	2
B09	4	1,525	0,006	0,98	0,98	0,48	0,40	2
B09	5	1,038	0,006	0,83	0,77	0,59	0,47	2
B09	6	2,417	0,006	1,00	1,06	0,46	0,29	3
B09	7	0,247	0,006	1,14	1,22	0,33	0,58	2
B09	8	1,158	0,006	1,00	0,99	0,47	0,45	2
B10	1	0,622	0,006	1,09	1,13	0,38	0,52	2
B10	2	1,040	0,006	1,17	1,26	0,33	0,47	2
B10	3	1,277	0,006	1,05	1,07	0,43	0,43	2
B10	4	0,725	0,006	0,85	0,80	0,56	0,51	2

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B10	5	0,953	0,006	0,99	0,99	0,46	0,48	2
B10	6	-	-	-	-	-	-	-
B10	7	0,066	0,006	1,17	1,23	0,30	0,60	1
B10	8	1,614	0,006	1,18	1,25	0,33	0,39	2
B11	1	-0,290	0,006	1,02	1,08	0,39	0,65	1
B11	2	-0,452	0,006	0,84	0,77	0,53	0,67	1
B11	3	0,601	0,006	0,84	0,80	0,57	0,53	2
B11	4	0,860	0,006	1,03	1,04	0,44	0,49	2
B11	5	2,168	0,006	0,95	0,95	0,50	0,32	3
B11	6	0,710	0,006	1,07	1,08	0,41	0,51	2
B11	7	0,924	0,006	0,94	0,92	0,50	0,48	2
B11	8	0,300	0,006	1,18	1,27	0,30	0,57	2
B12	1	-1,471	0,007	0,93	0,89	0,40	0,79	1
B12	2	0,087	0,006	0,89	0,83	0,52	0,60	1
B12	3	0,660	0,006	0,87	0,84	0,55	0,52	2
B12	4	0,817	0,006	0,79	0,73	0,61	0,50	2
B12	5	1,372	0,006	1,25	1,36	0,28	0,42	2
B12	6	2,502	0,006	1,01	1,04	0,45	0,28	3
B12	7	1,015	0,006	0,90	0,86	0,54	0,47	2
B12	8	0,490	0,006	0,93	0,89	0,50	0,54	2

Tabla 4.6 Dificultad y ajuste de los ítems al modelo Rasch de la prueba muestral de Lectura de 4.º grado de primaria

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B01	1	-2,991	0,012	0,72	0,49	0,32	0,95	1
B01	2	-1,668	0,008	0,87	0,80	0,39	0,83	2
B01	3	-1,804	0,008	0,82	0,72	0,34	0,87	2
B01	4	0,096	0,006	1,13	1,19	0,34	0,51	3
B01	5	-1,475	0,008	1,08	1,22	0,29	0,78	2
B01	6	-1,113	0,024	1,02	1,15	0,36	0,73	2
B01	7	-1,364	0,007	0,88	0,80	0,48	0,77	2
B01	8	-0,224	0,006	1,25	1,33	0,23	0,57	3
B01	9	-0,647	0,007	0,97	0,95	0,44	0,65	3
B01	10	-0,108	0,006	0,96	0,95	0,47	0,54	3
B01	11	-0,375	0,006	0,91	0,87	0,50	0,60	3
B01	12	-	-	-	-	-	-	-
B01	13	0,343	0,006	1,08	1,13	0,37	0,45	4
B01	14	-1,460	0,008	0,73	0,61	0,48	0,82	2
B01	15	0,614	0,007	1,15	1,24	0,34	0,45	4
B01	16	0,197	0,006	1,00	1,00	0,44	0,48	4
B01	17	-1,293	0,007	0,97	0,96	0,39	0,76	2
B01	18	-1,016	0,007	0,88	0,86	0,42	0,76	2
B02	1	1,316	0,012	1,06	1,27	0,35	0,28	4

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B02	2	1,089	0,012	1,14	1,33	0,30	0,31	4
B02	3	1,174	0,012	1,00	1,17	0,40	0,30	4
B02	4	-0,347	0,011	0,91	0,86	0,50	0,59	3
B02	5	-	-	-	-	-	-	-
B02	6	-1,274	0,013	1,00	1,01	0,37	0,76	2
B03	1	-1,736	0,014	0,86	0,76	0,47	0,82	2
B03	2	-2,374	0,017	0,86	0,70	0,42	0,89	1
B03	3	0,592	0,011	1,04	1,10	0,39	0,40	4
B03	4	-2,410	0,017	0,91	0,80	0,38	0,89	1
B03	5	-1,004	0,012	0,90	0,83	0,49	0,71	2
B03	6	-2,059	0,015	0,86	0,72	0,45	0,86	2
B04	1	0,045	0,011	1,02	1,02	0,43	0,51	3
B04	2	0,255	0,011	1,00	1,01	0,45	0,47	4
B04	3	-0,002	0,011	0,84	0,80	0,57	0,52	3
B04	4	-0,797	0,012	0,81	0,71	0,58	0,68	3
B04	5	1,165	0,012	1,11	1,32	0,32	0,30	4
B04	6	0,146	0,011	1,06	1,10	0,40	0,49	4
B05	1	-	-	-	-	-	-	-
B05	2	-0,749	0,007	0,96	0,90	0,46	0,66	3
B05	3	0,591	0,006	1,15	1,26	0,31	0,40	4
B05	4	-	-	-	-	-	-	-
B05	5	-0,287	0,006	1,02	1,01	0,42	0,58	3
B05	6	-0,037	0,030	0,97	0,96	0,46	0,53	3
B05	7	-0,877	0,007	0,94	0,91	0,46	0,69	3
B05	8	-1,454	0,007	0,90	0,78	0,46	0,78	2
B05	9	-0,829	0,007	0,98	0,94	0,44	0,68	3
B05	10	-0,439	0,006	0,91	0,88	0,50	0,61	3
B05	11	0,325	0,006	1,15	1,22	0,31	0,46	4
B05	12	0,769	0,031	0,87	0,79	0,55	0,37	4
B05	13	-0,219	0,006	0,97	0,95	0,47	0,54	3
B05	14	-0,832	0,007	1,00	1,02	0,41	0,68	3
B05	15	1,257	0,007	1,06	1,23	0,32	0,26	4
B05	16	-0,903	0,007	1,04	1,00	0,43	0,66	3
B05	17	-0,930	0,007	1,07	1,09	0,46	0,63	2
B05	18	-0,836	0,007	1,03	1,05	0,37	0,69	3
B06	1	-1,281	0,013	0,90	0,83	0,46	0,76	2
B06	2	-1,408	0,013	0,90	0,78	0,46	0,77	2
B06	3	-1,183	0,012	0,99	0,92	0,40	0,74	2
B06	4	-1,701	0,014	0,91	0,78	0,43	0,81	2
B06	5	-0,531	0,011	0,86	0,79	0,54	0,63	3
B06	6	-1,063	0,012	0,90	0,80	0,49	0,72	2
B07	1	0,927	0,012	1,00	1,15	0,34	0,29	4
B07	2	0,195	0,011	1,03	1,06	0,40	0,43	4
B07	3	-0,650	0,011	1,17	1,32	0,36	0,57	3
B07	4	-0,114	0,011	1,02	1,02	0,43	0,50	3
B07	5	0,584	0,011	0,99	1,04	0,43	0,40	4

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B07	6	-	-	-	-	-	-	-
B08	1	-1,760	0,014	0,82	0,62	0,51	0,82	2
B08	2	-1,102	0,012	0,91	0,85	0,47	0,73	2
B08	3	-0,461	0,011	0,93	0,90	0,49	0,61	3
B08	4	-0,752	0,012	0,84	0,76	0,55	0,67	3
B08	5	0,078	0,011	1,00	1,00	0,45	0,51	3
B08	6	0,041	0,011	1,19	1,25	0,29	0,51	3

Tabla 4.7 Dificultad y ajuste de los ítems al modelo Rasch de la prueba muestral de Matemática de 4.º grado de primaria

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B01	1	-2,040	0,016	1,05	1,15	0,34	0,86	1
B01	2	-2,407	0,017	0,94	0,91	0,39	0,89	1
B01	3	-	-	-	-	-	-	-
B01	4	-1,546	0,014	0,90	0,78	0,50	0,79	2
B01	5	0,993	0,012	1,09	1,17	0,43	0,37	4
B02	1	-	-	-	-	-	-	-
B02	2	0,039	0,012	0,99	1,05	0,49	0,55	3
B02	3	-1,616	0,014	0,94	0,95	0,41	0,82	2
B02	4	0,274	0,012	1,17	1,29	0,38	0,51	3
B02	5	-0,951	0,013	1,06	1,10	0,47	0,68	3
B03	1	-1,956	0,015	1,03	1,10	0,36	0,85	2
B03	2	1,244	0,012	1,12	1,35	0,40	0,33	4
B03	3	0,299	0,012	1,01	1,05	0,49	0,50	3
B03	4	-0,473	0,012	0,95	0,95	0,57	0,57	3
B03	5	-3,051	0,021	1,17	1,18	0,36	0,91	1
B04	1	-2,143	0,016	0,83	0,70	0,45	0,87	1
B04	2	-	-	-	-	-	-	-
B04	3	-0,140	0,012	1,05	1,13	0,47	0,54	3
B04	4	-1,102	0,013	1,09	1,12	0,43	0,72	2
B04	5	-0,209	0,012	1,02	1,07	0,48	0,59	3
B05	1	-2,414	0,017	0,90	0,86	0,40	0,89	1
B05	2	-0,815	0,012	1,09	1,24	0,40	0,70	3
B05	3	0,895	0,012	0,91	0,97	0,54	0,39	4
B05	4	-0,723	0,012	1,00	1,03	0,46	0,68	3
B05	5	-0,586	0,012	1,01	1,04	0,47	0,66	3
B06	1	-2,239	0,016	1,02	1,16	0,34	0,88	1
B06	2	-1,354	0,013	1,01	1,20	0,41	0,78	2
B06	3	-1,747	0,015	1,01	0,88	0,49	0,80	2
B06	4	-0,462	0,012	1,01	0,99	0,50	0,61	3
B06	5	-	-	-	-	-	-	-
B07	1	-2,791	0,011	0,99	1,28	0,31	0,92	1
B07	2	-1,624	0,008	0,73	0,57	0,48	0,85	2

Bloque	Posición en el bloque	Medida	Error	Infit	Outfit	ptme	p	Nivel
B08	1	-1,246	0,008	1,12	1,09	0,48	0,71	2
B08	2	0,700	0,007	1,06	1,17	0,45	0,42	4
B08	3	-1,593	0,008	0,88	0,82	0,38	0,84	2
B08	4	0,401	0,007	1,01	1,06	0,49	0,48	4
B08	5	1,104	0,007	1,05	1,22	0,45	0,35	4
B08	6	-0,036	0,007	1,19	1,30	0,36	0,59	3
B08	7	-0,396	0,007	1,16	1,23	0,38	0,63	3
B09	1	-2,225	0,009	0,89	0,73	0,43	0,87	1
B09	2	-0,201	0,007	0,89	0,87	0,56	0,59	3
B10	1	-	-	-	-	-	-	-
B10	2	-1,729	0,008	1,07	1,14	0,42	0,80	2
B10	3	0,767	0,007	1,14	1,23	0,41	0,41	4
B10	4	-	-	-	-	-	-	-
B10	5	-0,647	0,007	0,98	0,91	0,49	0,67	3
B10	6	0,309	0,007	0,99	1,02	0,50	0,50	3
B10	7	-1,231	0,008	0,85	0,77	0,55	0,75	2
B11	1	-1,889	0,009	0,82	0,71	0,40	0,87	2
B11	2	-0,250	0,007	1,03	1,02	0,48	0,59	3
B12	1	-1,891	0,009	0,93	0,88	0,43	0,84	2
B12	2	1,803	0,008	1,02	1,30	0,43	0,24	4
B12	3	0,590	0,007	1,05	1,11	0,47	0,44	4
B12	4	0,994	0,007	0,95	1,00	0,53	0,39	4
B12	5	0,204	0,007	1,07	1,14	0,45	0,52	3
B12	6	-1,506	0,008	1,03	0,98	0,42	0,79	2
B12	7	-2,724	0,011	0,97	0,84	0,40	0,90	1
B13	1	-2,086	0,009	1,05	1,06	0,44	0,84	1
B13	2	0,902	0,007	1,03	1,07	0,48	0,39	4
B14	1	-	-	-	-	-	-	-
B14	2	-1,351	0,008	0,86	0,74	0,45	0,81	2
B14	3	0,755	0,007	1,15	1,25	0,41	0,44	4
B14	4	1,203	0,007	1,03	1,23	0,45	0,34	4
B14	5	-0,682	0,007	0,85	0,76	0,54	0,70	3
B14	6	-1,775	0,008	0,93	0,89	0,46	0,82	2
B14	7	-2,973	0,012	0,98	0,81	0,36	0,92	1
B15	1	1,387	0,024	0,96	0,90	0,51	0,31	4
B16	1	2,171	0,027	0,89	0,77	0,51	0,20	4
B17	1	0,215	0,031	0,88	0,83	0,58	0,52	3
B18	1_1	1,011	0,028	1,17	1,24	0,46	0,42	4
B18	1_2	3,604	0,028	1,17	1,24	0,46	0,42	4

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

Wilson (2005) señala que la evidencia de validez basada en la estructura interna requiere, en primer lugar, que se tenga claridad conceptual sobre dicha estructura, dado el constructo de interés. En este sentido, un tipo de evidencia está relacionado con la unidimensionalidad de las medidas derivadas de la aplicación del test. 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 será 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. 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 ál., 1991). 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 el análisis factorial, que debe ser diferenciado del análisis de componentes principales (Preacher y MacCallum, 2003). El primero de ellos trata de describir la estructura latente de un conjunto de variables, usando factores que expliquen la varianza común; el segundo trata de reducir las variables a un conjunto menor de componentes enfocándose en la varianza total (Park, Dailey y Lemus, 2002; Preacher y MacCallum, 2003).

En el caso de las pruebas aplicadas en el 2019, se realizó un análisis de componentes principales de los residuos estandarizados luego de ajustar el modelo Rasch unidimensional (Linacre, 2019). El análisis factorial de los residuos sirve para detectar cualquier varianza que queda en los datos luego de extraer la dimensión principal (Bond y Fox, 2015); es decir, trata de modelar la varianza que no es explicada por el constructo de interés que pretende medir un test.

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 prefirió el análisis de componentes principales por su rigurosa base matemática.

Tabla 4.8 Análisis de unidimensionalidad de las medidas derivadas de la aplicación de las pruebas censales en secundaria

Prueba	Varianza de la dimensión principal (%)	Varianza modelada (%)	Primer autovalor	Primer autovalor (%)
Lectura, 2.º grado de secundaria	26,2	26,4	1,78	1,4
Matemática, 2.º grado de secundaria	33,4	33,1	1,49	1,4
Ciencia y Tecnología, 2.º grado de secundaria	22,4	22,4	2,02	1,7

Tabla 4.9 Análisis de unidimensionalidad de las medidas derivadas de la aplicación de las pruebas muestrales de primaria

Prueba	Varianza de la dimensión principal (%)	Varianza modelada (%)	Primer autovalor	Primer autovalor (%)
Lectura, 2.º grado de primaria	25,6	25,9	2,33	1,9
Matemática, 2.º grado de primaria	27,1	26,8	2,15	1,8
Lectura, 4.º grado de primaria	27,4	26,9	1,81	2,0
Matemática, 4.º grado de primaria	36,3	36,7	1,62	1,6

Linacre (2019) 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, no hay serias evidencias que atenten contra el supuesto de unidimensionalidad y, por lo 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 para las pruebas analizadas; es decir, no hay evidencias contundentes en contra del supuesto de unidimensionalidad del conjunto de ítems incluidos en estos instrumentos de medición.

4.4.3 Confiabilidad y consistencia de la clasificación

El índice de confiabilidad de la separación de las personas (R_p) es el coeficiente utilizado para las pruebas aplicadas de la UMC, el cual es análogo al de consistencia interna alpha de Cronbach, pero produce mejores estimaciones, pues los valores numéricos son lineales si los datos se ajustan al modelo Rasch aplicado. 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 de 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 y Fox, 2015).

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 bastante variabilidad; asimismo, observó valores propuestos con los diversos métodos para obtener la confiabilidad, que 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, que se refieren a la dispersión de los datos medidos como el número de errores estándar que separan a las personas (Schumacker y Smith, 2007). El índice de separación de personas (G_p) representa la variabilidad ajustada de las personas dividida entre el error estándar de medición promedio.

Tabla 4.10 Análisis de confiabilidad de las medidas derivadas de la aplicación de las pruebas censales en secundaria

Prueba	R_p	G_p
Lectura, 2.º grado de secundaria	0,87	2,54
Matemática, 2.º grado de secundaria	0,87	2,56
Ciencia y Tecnología, 2.º grado de secundaria	0,68	1,45

Considerando el valor de R_p de la prueba censal aplicada en el 2019 para 2.º grado de secundaria, se puede apreciar que la varianza de error en la prueba de Lectura es de 13 %, en la de Matemática es de 13 % y en Ciencia y Tecnología, de 32 %. Por lo tanto, es posible afirmar que las medidas derivadas de aplicar dichas pruebas poseen adecuadas evidencias de confiabilidad.

Tabla 4.11 Análisis de confiabilidad de las medidas derivadas de la aplicación de las pruebas muestrales en primaria

Prueba	R_p	G_p
Lectura, 2.º grado de primaria	0,89	2,79
Matemática, 2.º grado de primaria	0,90	2,97
Lectura, 4.º grado de primaria	0,86	2,53
Matemática, 4.º grado de primaria	0,89	2,79

Del mismo modo, en las pruebas muestrales aplicadas en el 2019, se puede apreciar que la varianza de error como máximo es de 14 % en la prueba de Lectura de 4.º grado de primaria y como mínimo de 10 % en la de Matemática de 2.º grado de primaria, por lo cual es posible afirmar que las medidas derivadas de aplicar dichas pruebas poseen adecuadas evidencias de confiabilidad.

Otros indicadores importantes para el análisis son los de precisión y consistencia de clasificación. La consistencia de la clasificación es el grado de acuerdo en dos administraciones independientes o paralelas de un instrumento de medición. Se espera que el alumno sea clasificado en la misma categoría al repetirse la evaluación. Por otro lado, la precisión de la clasificación implica el grado en el cual

la clasificación observada coincide con la clasificación verdadera (Kim, Choi, Um y Kim, 2006).

En la práctica, es difícil lograrlo; por ello, se han propuesto métodos para estimarla a partir de una sola aplicación. En las evaluaciones del 2019, se utilizó el método de Rudner, que es 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.

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

	Lectura		Matemática		Ciencia y Tecnología	
	Precisión	Consistencia	Precisión	Consistencia	Precisión	Consistencia
Nacional	0,74	0,65	0,75	0,66	0,68	0,56
Hombre	0,74	0,65	0,74	0,66	0,68	0,56
Mujer	0,74	0,65	0,75	0,66	0,68	0,56
Urbano	0,74	0,64	0,74	0,65	0,67	0,56
Rural	0,77	0,69	0,81	0,74	0,69	0,58
Estatad	0,74	0,65	0,75	0,67	0,68	0,56
No estadad	0,74	0,64	0,73	0,64	0,67	0,56

Tabla 4.13 Indicadores del análisis de precisión y consistencia de la clasificación de las pruebas muestrales de 2.º grado de primaria

	Lectura		Matemática	
	Precisión	Consistencia	Precisión	Consistencia
Nacional	0,87	0,82	0,90	0,86
Hombre	0,87	0,82	0,90	0,86
Mujer	0,87	0,82	0,90	0,86
Urbano	0,88	0,83	0,90	0,86
Rural	0,86	0,80	0,92	0,89
Estatad	0,87	0,82	0,90	0,86
No estadad	0,88	0,83	0,90	0,86
Polidocente completo	0,88	0,82	0,90	0,86
Unidocente / Multigrado	0,86	0,81	0,92	0,89

Tabla 4.14 Indicadores del análisis de precisión y consistencia de la clasificación de las pruebas muestrales de 4.º grado de primaria

	Lectura		Matemática	
	Precisión	Consistencia	Precisión	Consistencia
Nacional	0,78	0,69	0,80	0,72
Hombre	0,78	0,69	0,81	0,73
Mujer	0,78	0,70	0,80	0,72
Urbano	0,78	0,70	0,81	0,73
Rural	0,77	0,68	0,79	0,70
Estatal	0,78	0,69	0,80	0,72
No estatal	0,80	0,72	0,81	0,74
Polidocente completo	0,78	0,70	0,81	0,73
Unidocente / Multigrado	0,77	0,68	0,79	0,70

En general, las pruebas mantienen buenos indicadores de precisión y consistencia de la clasificación de los niveles de logro.

4.4.4 Equiparación de medidas

En el marco del análisis Rasch, las puntuaciones del rasgo latente tienen un origen y una escala de medición arbitrarios (Bond y Fox, 2015). Por ello, dos instrumentos que miden lo mismo, aplicados a distintas muestras de personas, no se encontrarán en la misma métrica. A fin de establecer las comparaciones necesarias, se deben poner ambas pruebas en la misma escala o equipararlas. Esta conversión, en ocasiones, implica el uso de una constante aditiva y, en otras, el uso de una constante multiplicativa, además de la aditiva (Livingston, 2004).

Diversas síntesis teóricas que tratan sobre la equiparación de puntuaciones (Kolen y Brennan, 2004; Navas, 1996) citan los trabajos de Angoff y Lord (ambos en la primera mitad de los años ochenta) como los pioneros en la equiparación. 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, se tienen dos test (X e Y) y se desea equiparar el test X con el test Y . Por lo tanto, el objetivo es encontrar un sistema o función para convertir la métrica de X en Y .

Revisando la literatura sobre el tema (Ho y Osborn, 2005; Kolen y Brennan, 2004; Livingston, 2004; Navas, 1996; Zhu, 1998), se puede señalar que hay 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 censales de 2.º grado de

primaria en la misma métrica que las pruebas aplicadas en el 2007 por la UMC. En este diseño, se trabaja con dos muestras de personas que no necesariamente han sido extraídas de la misma población. En cada muestra, se aplica una única forma del test. Lo esencial de este método es que en cada grupo se administra un conjunto de ítems o un test común que permite establecer la equivalencia entre los test que se quieren equiparar.

Una vez obtenidos los datos, se utilizó 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, 2004), e implica el cálculo de dos constantes de equiparación *a* y *b*, de tal manera que surge la siguiente fórmula:

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

donde:

$$a = \frac{S_{medida_{anterior}}}{S_{medida_{actual}}}$$

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

Un requisito importante para realizar la comparación del rendimiento entre dos o más grupos es asegurar la invarianza de la dificultad de los ítems de la prueba a lo largo del tiempo (Kolen y Brennan, 2004). Esto supone que el grado de dificultad de cada ítem debe mantenerse constante⁸, 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 este tipo de análisis según la metodología propuesta por Bond y Fox (2015). De esta manera, solo se utilizan en la equiparación aquellos ítems sin funcionamiento diferencial a lo largo del tiempo.

Desde la tabla 4.15 a la 4.21 se presentan los resultados de la equiparación por ítems comunes de las pruebas aplicadas en las evaluaciones del 2019.

Tabla 4.15 Análisis FDI de las medidas de la prueba censal de Lectura de 2.º grado de secundaria usadas en la equiparación de 2018 con 2019

Bloque	Posición en el bloque	2018		2019		Eqp	FDI	
		Medida	Error	Medida	Error	Medida Trans	Medida	Error
B01	14	2,397	0,005	1,045	0,003	2,065	0,332	0,006
B01	17	0,581	0,005	-0,749	0,003	0,375	0,206	0,006
B01	16	1,315	0,005	0,227	0,003	1,294	0,021	0,006
B01	18	1,012	0,005	-0,350	0,003	0,751	0,261	0,006
B01	19	0,332	0,005	-1,098	0,004	0,046	0,286	0,006

⁸En realidad, la dificultad del ítem debe mantenerse dentro de ciertos márgenes de error al comparar su dificultad en dos poblaciones diferentes.

Bloque	Posición en el bloque	2018		2019		Eqp	FDI	
		Medida	Error	Medida	Error	Medida Trans	Medida	Error
B03	1	0,064	0,006	-0,606	0,007	0,509	-0,445	0,009
B03	2	1,677	0,005	0,767	0,007	1,803	-0,126	0,009
B03	3	0,137	0,006	-0,629	0,007	0,488	-0,351	0,009
B03	4	0,528	0,005	-0,461	0,007	0,646	-0,118	0,009
B03	5	1,481	0,005	0,715	0,007	1,754	-0,273	0,009
B01	1	-1,807	0,007	-3,173	0,007	-1,910	0,103	0,010
B01	2	0,438	0,003	-0,760	0,003	0,364	0,074	0,004
B01	3	-0,993	0,005	-2,325	0,005	-1,110	0,117	0,007
B01	4	-0,747	0,004	-2,064	0,004	-0,865	0,118	0,006
B01	5	0,866	0,003	-0,375	0,003	0,727	0,139	0,004
B01	6	1,649	0,003	0,545	0,003	1,594	0,055	0,004
B01	7	2,701	0,033	1,638	0,033	2,624	0,077	0,047
B05	1	2,140	0,005	0,945	0,007	1,971	0,169	0,009
B05	2	2,722	0,006	1,792	0,009	2,769	-0,047	0,011
B05	3	1,984	0,005	0,860	0,007	1,891	0,093	0,009
B05	4	2,041	0,005	0,952	0,007	1,978	0,063	0,009
B05	5	2,251	0,005	1,162	0,008	2,175	0,076	0,009
B05	6	2,037	0,005	1,121	0,008	2,137	-0,100	0,009
B07	1	1,139	0,005	0,110	0,007	1,184	-0,045	0,009
B07	2	-0,585	0,007	-1,543	0,009	-0,374	-0,211	0,011
B07	3	1,026	0,005	-0,157	0,007	0,932	0,094	0,009
B07	4	1,286	0,005	0,437	0,007	1,492	-0,206	0,009
B07	5	1,790	0,005	0,870	0,007	1,900	-0,110	0,009
B07	6	1,575	0,005	0,634	0,007	1,678	-0,103	0,009
B03	6	1,129	0,005	0,208	0,007	1,276	-0,147	0,009

Tabla 4.16 Análisis FDI de las medidas de la prueba censal de Matemática de 2.º grado de secundaria usadas en la equiparación de 2018 con 2019

Bloque	Posición en el bloque	2018		2019		Eqp	FDI	
		Medida	Error	Medida	Error	Medida Trans	Medida	Error
B01	5	0,735	0,003	0,217	0,007	1,117	-0,382	0,008
B05	4	1,261	0,007	0,406	0,007	1,298	-0,037	0,010
B13	2	0,726	0,003	-0,383	0,003	0,544	0,182	0,004
B07	3	2,358	0,003	1,214	0,008	2,070	0,288	0,009
B12	1	0,978	0,007	-0,053	0,003	0,860	0,118	0,008
B18	5	2,045	0,007	1,346	0,003	2,196	-0,151	0,008
B12	6	-1,050	0,005	-1,863	0,005	-0,869	-0,181	0,007
B09	2	1,997	0,007	1,161	0,008	2,019	-0,022	0,011
B11	1	-0,550	0,009	-1,720	0,004	-0,732	0,182	0,010
B09	3	2,062	0,007	1,422	0,008	2,268	-0,206	0,011
B15	1	-0,033	0,004	-1,162	0,004	-0,199	0,166	0,006
B16	3	0,719	0,007	-0,210	0,003	0,710	0,009	0,008
B02	5	0,699	0,007	-0,187	0,007	0,732	-0,033	0,010
B05	1	-0,275	0,004	-1,188	0,009	-0,224	-0,051	0,010

Bloque	Posición en el bloque	2018		2019		Eqp	FDI	
		Medida	Error	Medida	Error	Medida Trans	Medida	Error
B10	5	1,193	0,003	0,372	0,007	1,265	-0,072	0,008
B16	4	2,021	0,007	1,038	0,003	1,901	0,120	0,008
B12	3	1,130	0,007	0,249	0,003	1,148	-0,018	0,008
B18	3	-0,137	0,004	-1,316	0,004	-0,346	0,209	0,006
B01	3	2,026	0,007	0,985	0,007	1,851	0,175	0,010
B14	1	-1,394	0,005	-2,265	0,005	-1,253	-0,141	0,007
B04	3	0,483	0,007	-0,318	0,008	0,607	-0,124	0,011
B08	4	1,277	0,007	0,512	0,007	1,399	-0,122	0,010
B09	1	0,853	0,003	0,081	0,007	0,988	-0,135	0,008
B06	1	-0,424	0,009	-1,512	0,009	-0,534	0,110	0,013
B11	2	-0,394	0,004	-1,258	0,004	-0,291	-0,103	0,006
B16	2	0,654	0,007	-0,571	0,003	0,365	0,289	0,008
B12	4	2,526	0,008	1,767	0,004	2,598	-0,072	0,009
B04	2	0,201	0,008	-0,785	0,008	0,161	0,040	0,011
B05	5	1,144	0,007	0,197	0,007	1,098	0,046	0,010
B14	5	-0,661	0,009	-1,542	0,004	-0,562	-0,099	0,010
B12	2	0,282	0,003	-0,671	0,004	0,269	0,013	0,005

Tabla 4.17 Análisis FDI de las medidas de la prueba censal de Ciencia y Tecnología de 2.º grado de secundaria usadas en la equiparación de 2018 con 2019

Bloque	Posición en el bloque	2018		2019		Eqp	FDI	
		Medida	Error	Medida	Error	Medida Trans	Medida	Error
B06	9	-0,163	0,005	-0,185	0,005	-0,193	0,030	0,007
B06	12	-0,201	0,005	-0,104	0,005	-0,111	-0,090	0,007
B06	13	1,072	0,006	1,031	0,006	1,039	0,033	0,008
B01	3	-0,819	0,005	-0,657	0,005	-0,672	-0,147	0,007
B01	5	-0,341	0,005	-0,322	0,005	-0,332	-0,009	0,007
B01	4	1,398	0,006	1,400	0,006	1,412	-0,014	0,008
B02	13	0,074	0,005	0,141	0,005	0,137	-0,063	0,007
B02	14	0,089	0,005	0,034	0,005	0,029	0,060	0,007
B04	9	0,692	0,005	0,732	0,005	0,736	-0,044	0,007
B02	9	1,513	0,006	1,415	0,006	1,428	0,085	0,008
B02	10	0,454	0,005	0,314	0,005	0,312	0,142	0,007
B04	2	0,250	0,005	0,176	0,005	0,172	0,078	0,007
B04	3	0,056	0,005	0,044	0,005	0,039	0,017	0,007
B03	4	0,062	0,005	0,136	0,005	0,132	-0,070	0,007
B05	10	0,963	0,005	0,966	0,006	0,973	-0,010	0,008
B04	4	1,513	0,006	1,578	0,006	1,593	-0,080	0,008
B03	7	0,407	0,005	0,470	0,005	0,470	-0,063	0,007
B03	2	0,705	0,005	0,843	0,006	0,848	-0,143	0,008
B05	1	-1,164	0,006	-1,271	0,006	-1,294	0,130	0,008
B05	2	-0,118	0,005	-0,229	0,005	-0,238	0,120	0,007
B03	10	0,772	0,005	0,704	0,005	0,707	0,065	0,007
B06	2	0,535	0,005	0,427	0,005	0,427	0,108	0,007

Bloque	Posición en el bloque	2018		2019		Eqp	FDI	
		Medida	Error	Medida	Error	Medida Trans	Medida	Error
B06	4	-2,125	0,008	-1,992	0,008	-2,024	-0,101	0,011
B05	13	-1,179	0,006	-1,115	0,006	-1,136	-0,043	0,008
B05	14	1,058	0,006	1,177	0,006	1,187	-0,129	0,008
B04	8	0,906	0,050	0,969	0,050	0,976	-0,070	0,071
B03	8	-0,407	0,005	-0,343	0,005	-0,353	-0,054	0,007
B03	9	0,509	0,005	0,488	0,005	0,488	0,021	0,007
B01	16	-0,609	0,005	-0,293	0,005	-0,303	-0,306	0,007
B05	5	0,404	0,005	0,364	0,005	0,363	0,041	0,007
B05	6	0,314	0,005	0,402	0,005	0,401	-0,087	0,007
B05	7	-0,097	0,005	-0,036	0,005	-0,042	-0,055	0,007
B05	16	-0,418	0,005	-0,322	0,005	-0,332	-0,086	0,007
B03	5	0,083	0,005	0,050	0,005	0,045	0,038	0,007
B01	1	-0,122	0,005	-0,191	0,005	-0,199	0,077	0,007
B01	2	-1,059	0,006	-1,095	0,006	-1,115	0,056	0,008
B03	14	-1,217	0,006	-1,186	0,006	-1,207	-0,010	0,008
B03	15	-0,274	0,005	-0,247	0,005	-0,256	-0,018	0,007
B03	6	-0,295	0,005	-0,178	0,005	-0,186	-0,109	0,007
B04	5	0,748	0,005	0,743	0,005	0,747	0,001	0,007
B06	3	0,336	0,005	0,305	0,005	0,303	0,033	0,007
B04	7	0,053	0,005	0,105	0,005	0,100	-0,047	0,007
B04	13	-0,339	0,005	-0,421	0,005	-0,432	0,093	0,007
B04	14	0,640	0,005	0,676	0,005	0,679	-0,039	0,007
B06	7	0,707	0,005	0,792	0,005	0,796	-0,089	0,007
B02	1	1,177	0,006	1,141	0,006	1,150	0,027	0,008
B02	2	0,221	0,005	0,166	0,005	0,162	0,059	0,007
B01	9	1,348	0,006	1,374	0,006	1,386	-0,038	0,008
B01	10	1,208	0,006	1,217	0,006	1,227	-0,019	0,008
B04	12	-2,578	0,009	-2,342	0,008	-2,379	-0,199	0,012
B01	11	-0,289	0,005	-0,323	0,005	-0,333	0,044	0,007
B04	6	0,666	0,005	0,671	0,005	0,674	-0,008	0,007
B06	1	-0,591	0,005	-0,660	0,006	-0,675	0,084	0,008
B02	6	-1,871	0,007	-1,901	0,007	-1,932	0,061	0,010
B04	11	-0,105	0,005	-0,068	0,005	-0,075	-0,030	0,007
B06	10	-1,672	0,007	-1,550	0,007	-1,576	-0,096	0,010
B06	11	-1,386	0,006	-1,427	0,006	-1,452	0,066	0,008
B02	8	-0,357	0,005	-0,325	0,005	-0,335	-0,022	0,007
B03	1	0,412	0,005	0,302	0,005	0,300	0,112	0,007
B04	1	-0,846	0,005	-0,999	0,006	-1,018	0,172	0,008
B05	12	-0,032	0,005	0,004	0,005	-0,002	-0,030	0,007
B03	3	-0,254	0,005	-0,293	0,005	-0,303	0,049	0,007
B01	12	-0,067	0,005	-0,215	0,005	-0,224	0,157	0,007
B04	10	-0,351	0,005	-0,674	0,005	-0,689	0,338	0,007
B02	7	-1,456	0,006	-1,531	0,006	-1,557	0,101	0,008
B06	15	-2,229	0,008	-2,110	0,008	-2,144	-0,085	0,011
B06	16	-1,681	0,007	-1,692	0,007	-1,720	0,039	0,010
B06	5	0,354	0,005	0,389	0,005	0,388	-0,034	0,007

Tabla 4.18 Análisis FDI de las medidas de la prueba muestral de Lectura de 2.º grado de primaria usadas en la equiparación 2018 con 2019

Bloque	Posición en el bloque	2018		2019		Eqp	FDI	
		Medida	Error	Medida	Error	Medida Trans	Medida	Error
B01	1	-0,645	0,041	-2,535	0,010	-0,674	0,028	0,042
B01	3	-0,306	0,036	-2,120	0,009	-0,273	-0,033	0,037
B01	4	1,128	0,024	-0,681	0,006	1,115	0,013	0,025
B02	2	0,451	0,028	-1,431	0,007	0,392	0,059	0,029
B02	3	2,014	0,021	0,252	0,005	2,015	-0,001	0,022
B02	5	1,046	0,024	-0,610	0,006	1,184	-0,138	0,025
B02	6	0,895	0,025	-0,928	0,006	0,877	0,018	0,026
B02	7	0,827	0,025	-0,939	0,006	0,866	-0,039	0,026
B02	8	2,354	0,021	0,591	0,005	2,343	0,011	0,022
B02	9	1,984	0,021	0,200	0,005	1,965	0,019	0,022
B02	10	1,026	0,024	-0,753	0,006	1,046	-0,020	0,025
B02	4	1,841	0,022	-0,044	0,006	1,730	0,111	0,023
B02	1	0,436	0,028	-1,527	0,007	0,299	0,137	0,029
B16	1	1,716	0,031	-0,072	0,008	1,703	0,013	0,032
B16	2	2,681	0,031	0,986	0,008	2,724	-0,042	0,032
B16	3	2,377	0,030	0,615	0,008	2,366	0,011	0,031
B16	4	1,939	0,031	0,069	0,008	1,839	0,100	0,032
(B15,B19)	(1,1)	2,594	0,031	0,822	0,008	2,565	0,028	0,032
(B15,B19)	(2,2)	2,534	0,031	0,851	0,008	2,593	-0,059	0,032
(B15,B19)	(3,4)	1,637	0,031	-0,099	0,008	1,677	-0,039	0,032
B04	2	1,949	0,030	0,181	0,008	1,947	0,002	0,031
B04	1	0,676	0,037	-0,897	0,009	0,907	-0,231	0,038
B04	3	2,509	0,030	0,634	0,008	2,384	0,125	0,031
B04	5	2,589	0,031	0,793	0,008	2,537	0,052	0,032
B01	2	-0,535	0,039	-2,422	0,010	-0,565	0,029	0,040
(B15,B19)	(6,5)	3,192	0,033	1,433	0,008	3,155	0,038	0,034
B16	5	1,728	0,031	-0,051	0,008	1,723	0,005	0,032
(B14,B22)	(2,1)	2,391	0,030	0,682	0,008	2,430	-0,040	0,031
(B14,B22)	(3,3)	2,560	0,030	0,786	0,008	2,531	0,030	0,031
(B14,B22)	(4,4)	2,644	0,031	1,037	0,008	2,773	-0,129	0,032
(B03,B08)	(1,1)	1,850	0,031	0,110	0,008	1,878	-0,029	0,032
(B03,B08)	(2,3)	2,604	0,031	0,859	0,008	2,601	0,003	0,032
(B03,B08)	(3,4)	2,565	0,031	0,855	0,008	2,597	-0,032	0,032

Tabla 4.19 Análisis FDI de las medidas de la prueba muestral de Matemática de 2.º grado de primaria usadas en la equiparación 2018 con 2019

Bloque	Posición en el bloque	2018		2019		Eqp	FDI	
		Medida	Error	Medida	Error	Medida Trans	Medida	Error
B12	3	0,624	0,030	-0,072	0,006	0,726	-0,102	0,031
B09	5	0,953	0,031	0,200	0,006	1,094	-0,141	0,032
B12	4	1,475	0,022	0,041	0,006	0,879	0,596	0,023
B10	4	0,789	0,021	-0,025	0,006	0,790	-0,001	0,022
B09	2	-0,220	0,022	-0,758	0,006	-0,203	-0,017	0,023

Bloque	Posición en el bloque	2018		2019		Eqp	FDI	
		Medida	Error	Medida	Error	Medida Trans	Medida	Error
B11	1	-0,280	0,022	-0,755	0,006	-0,199	-0,081	0,023
B12	1	-1,621	0,025	-1,604	0,007	-1,348	-0,273	0,026
B11	7	1,122	0,022	0,118	0,006	0,983	0,139	0,023
B11	3	0,640	0,030	-0,114	0,006	0,669	-0,029	0,031
B09	4	1,480	0,022	0,550	0,006	1,568	-0,088	0,023
B12	7	1,390	0,022	0,183	0,006	1,071	0,319	0,023
B11	2	0,316	0,021	-0,871	0,006	-0,356	0,672	0,022
B09	1	-0,489	0,022	-1,150	0,006	-0,733	0,244	0,023
B12	2	0,064	0,021	-0,484	0,006	0,168	-0,104	0,022
B12	8	0,475	0,021	-0,194	0,006	0,561	-0,086	0,022
B10	1	0,439	0,021	-0,099	0,006	0,690	-0,251	0,022
B10	5	0,857	0,021	0,139	0,006	1,012	-0,155	0,022
B11	5	2,150	0,023	1,012	0,006	2,194	-0,044	0,024
B11	6	0,961	0,022	-0,036	0,006	0,775	0,186	0,023
B12	6	2,763	0,025	1,252	0,006	2,519	0,244	0,026
B10	7	0,631	0,030	-0,499	0,006	0,148	0,483	0,031
B10	8	1,468	0,031	0,614	0,006	1,655	-0,187	0,032
B11	8	0,495	0,030	-0,331	0,006	0,375	0,120	0,031
B10	3	0,956	0,030	0,372	0,006	1,327	-0,371	0,031
B09	6	2,226	0,034	1,191	0,006	2,436	-0,210	0,035
B12	5	1,304	0,031	0,440	0,006	1,419	-0,115	0,032
B09	7	0,144	0,030	-0,369	0,006	0,324	-0,180	0,031
B10	2	0,688	0,030	0,201	0,006	1,096	-0,408	0,031
B09	8	1,066	0,031	0,286	0,006	1,211	-0,145	0,032
B11	4	1,239	0,031	0,072	0,006	0,921	0,318	0,032
B09	3	0,308	0,030	-0,137	0,006	0,638	-0,330	0,031

Tabla 4.20 Constantes de equiparación para las pruebas aplicadas en las evaluaciones censales

Prueba	a	b
Lectura, 2.º grado de secundaria	0,933	1,096
Matemática, 2.º grado de secundaria	0,955	0,910
Ciencia y Tecnología, 2.º grado de secundaria	1,013	-0,006

Tabla 4.21 Constantes de equiparación para las pruebas aplicadas en las evaluaciones muestrales

Prueba	a	b
Lectura, 2.º grado de primaria	0,965	1,772
Matemática, 2.º grado de primaria	1,391	0,760

Nota. La equiparación de las pruebas aplicadas en los años 2018 y 2019 fue realizada con tres métodos distintos para verificar su consistencia: ítems en común con calibraciones por separado, calibración concurrente y anclaje de valores en Winsteps. En el caso de cuarto de primaria, solo dos de los métodos mostraban resultados convergentes. Por ello se descartó la equiparación por ítems en común con calibración por separado y se usó el anclaje de valores en Winsteps, que no produce constantes de equiparación. En los demás casos, los tres métodos de equiparación mostraban resultados similares.

4.5 Niveles de logro y presentación de resultados

El diseño de las evaluaciones permite estimar lo que saben y hacen los estudiantes a partir de su desempeño en las pruebas, con respecto a lo que deberían saber y hacer. De acuerdo con ello, la interpretación de los resultados de la ECE está referida a criterios; es decir, los resultados se interpretan considerando un criterio o estándar de logro (Muñiz, 1998). Para esto, 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. Ello tiene por intención sostener los mismos puntos de corte en las ediciones posteriores y asegurar que los resultados sean comparables en el tiempo. Este año, fue utilizado para determinar los niveles de la prueba de Ciencia y Tecnología.

Se utiliza el método *bookmark* (Cizek y Bunch, 2007; Karantonis y Sireci, 2006; Lewis, Mitzel y Green, 1996), que 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, consiste en colocar marcas (tantas como cortes se hayan preestablecido) en un cuadernillo de ítems ordenado por dificultad (estando esta dificultad estimada mediante procedimientos Rasch o de teoría de respuesta al ítem). La pregunta típica que guía el establecimiento de cortes es la siguiente: “¿Hasta qué ítem debe ser capaz de resolver un estudiante, como mínimo, para ser considerado parte del nivel?”. El procedimiento establece que los jueces, organizados en grupos pequeños,

determinan, en tres rondas, 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 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, se les entrega un reporte de discrepancias en el que se señala qué tan distintos son sus juicios con 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 2004* (Ministerio de Educación, 2005), se establece que un estudiante está en un nivel de logro si tiene una probabilidad mayor o igual a 0,62 para 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 sea medio *logit* más que la medida de dificultad de dicho ítem. Después de esta adición, estos valores son transformados a medida 500 (M500) y redondeados a dos decimales. Finalmente, se asigna a cada estudiante, de acuerdo a su M500, a un nivel de logro determinado.

En las tablas 4.22 y 4.23, se presentan los puntos de corte (en M500) de las evaluaciones censal y muestrales.

Tabla 4.22 Medidas (en M500) que marcan los puntos de corte de los alumnos en la evaluación censal de 2.º grado de secundaria

Área	<Nivel 1 vs. Nivel 1	Nivel 1 vs. Nivel 2	Nivel 2 vs. Nivel 3
Lectura, 2.º grado de secundaria	505,14	580,61	641,25
Matemática, 2.º grado de secundaria	519,67	595,96	649,38
Ciencia y Tecnología, 2.º grado de secundaria	374,60	509,58	628,48

Tabla 4.23 Medidas (en M500) que marcan los puntos de corte de los alumnos en las evaluaciones muestrales

Área	<Nivel 1 vs. Nivel 1	Nivel 1 vs. Nivel 2	Nivel 2 vs. Nivel 3
Lectura, 2.º grado de primaria	458,39	583,66	—
Matemática, 2.º grado de primaria	512,22	639,21	—
Lectura, 4.º grado de primaria	356,92	444,72	522,03
Matemática, 4.º grado de primaria	351,90	422,21	526,46

Finalmente, es importante señalar que las medidas de habilidad de los alumnos, 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 las tablas 4.24 y 4.25.

Tabla 4.24 Valores de a y b utilizados para la transformación lineal de las medidas de la prueba censal de 2.º grado de secundaria

Prueba	a	b
Lectura, 2.º grado de secundaria	474,896	73,414
Matemática, 2.º grado de secundaria	463,157	73,781
Ciencia y Tecnología, 2.º grado de secundaria	496,120	134,513

Tabla 4.25 Valores de a y b utilizados para la transformación lineal de las medidas de las pruebas muestrales

Prueba	a	b
Lectura, 2.º grado de primaria	423,614	63,585
Matemática, 2.º grado de primaria	468,321	72,440
Lectura, 4.º grado de primaria	474,896	73,414
Matemática, 4.º grado de primaria	463,157	73,781

Referencias

Referencias

- American Educational Research Association, American Psychological Association y National Council of Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bond, T. G., y Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human Sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Browning, D. (2003). *Common definitions: Adaptations, accommodations, modifications*. Recuperado de 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>
- Cizek, G. J. (1993). Reconsidering standards and criteria. *Journal of Educational Measurement*, 30(2), 93-106. <https://doi.org/10.1111/j.1745-3984.1993.tb01068.x>
- Cizek, G. J., y Bunch, M. B. (2007). *Standard settings*. California, CA: SAGE.
- Embretson, S., y Reise, S. (2000). *Item Response Theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fisher, W. (1992). Reliability statistics. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 2, 1996* (p. 238). Chicago, IL: MESA Press.
- Fox, C. (1999). An introduction to the Partial Credit model for developing nursing assessments. *Journal of Nursing Education*, 38(8), 340-346.
- Hambleton, R., Swaminathan, H. y Rogers, J. (1991). *Fundamentals of Item Response Theory*. Thousand Oaks, CA: SAGE.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139-164. <https://doi.org/10.1177/014662168500900204>
- Ho, C., y Osborn, S. E. (2005). Test equating by common items and common subjects: Concepts and applications. *Practical Assessment, Research & Evaluation*, 10(4), 1-19.
- Ingebo, G. (1997). *Probability in the measure of achievement*. Chicago, IL: MESA.
- Karantonis, A., y Sireci, S. G. (2006). The bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice*, 25(1), (4-12). <https://doi.org/10.1111/j.1745-3992.2006.00047.x>
- Kim, D., Choi, S., Um, K. y Kim, J. (2006, Abril). *A comparison of methods for estimating classification consistency*. Trabajo presentado en la Annual Meeting of the National Council on Education in Measurement, San Francisco, CA.

- Kolen, M., y Brennan, R. (2004). *Test equating, scaling and linking* (2nd ed.). New York, NY: Springer.
- 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. y Green, D. R. (1996, Junio). Standard setting: A bookmark approach. En D. R. Green (Ed.), *Irt-based standard-setting procedures utilizing behavioral anchoring*. Simposio presentado en el Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, 2(3), 266-283.
- Linacre, J. M. (2001). Category, step and threshold: Definitions & disordering. *Rasch Measurement Transactions*, 15(1), 794.
- Linacre, J. M. (2019). *Winsteps manual*. Recuperado de <https://www.winsteps.com/a/Winsteps-Manual.pdf>
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. New Jersey, NJ: Educational Testing Service.
- Lohr, S. (1999). *Sampling: Design and Analysis*. Pacific Grove, CA: Duxbury Press.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174. <https://doi.org/10.1007/BF02296272>
- Ministerio de Educación. (2005). *Evaluación nacional del rendimiento estudiantil 2004. Informe descriptivo de resultados*. Lima: Oficina de Medición de la Calidad de los Aprendizajes.
- Ministerio de Educación. (2009). *Evaluación Censal de Estudiantes (ECE). Segundo grado de primaria y cuarto grado de primaria de IE EIB. Marco de trabajo*. Lima: Oficina de Medición de la Calidad de los Aprendizajes.
- Ministerio de Educación. (2015). *Marco de fundamentación de las pruebas de rendimiento de la Evaluación Censal de Estudiantes de 2do grado de secundaria 2015*. Lima: Oficina de Medición de la Calidad de los Aprendizajes.
- Ministerio de Educación. (2019). *Reporte técnico de la Evaluación Muestral de Estudiantes (EM 2018) de 2.º grado de primaria*. Lima: Oficina de Medición de la Calidad de los Aprendizajes.
- Moreno, R., Martínez, R. J. y Muñiz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema*, 16(3), 490-497.
- Muñiz, J. (1998). *Teoría clásica de los tests* (2nd ed.). Madrid: Ediciones Pirámide.
- Navas, M. (1996). Equiparación de puntuaciones. En J. Muñiz (Ed.), *Psicometría* (pp. 293-370). Madrid: Universitat.
- Nunnally, J., y Bernstein, I. (1995). *Teoría psicométrica* (3rd ed.). México: McGraw-Hill.

- Park, H. S., Dailey, R. y 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>
- Preacher, K. J., y MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2(1), 13-43. https://doi.org/10.1207/S15328031US0201_02
- Schulz, E. M. (1990). Functional assessment of fit. En J. M. Linacre (Ed.), *Rasch Measurement Transactions*, 1990 (p. 82). Chicago, IL: MESA Press.
- Schumacker, R. E., y Smith, E. V. (2007). Reliability: A Rasch perspective. *Educational and Psychological Measurement*, 67(3), 394-409. <https://doi.org/10.1177/0013164406294776>
- Shepard, L. (1980). Standard setting issues and methods. *Applied Psychological Measurement*, 4(4), 447-467. <https://doi.org/10.1177/014662168000400403>
- Smith, R., y Kramer, G. (1989). Response pattern analysis with supplemental store reports. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1, 1995* (pp.33-35). Chicago, IL: MESA Press.
- Wilson, M. (2005). *Constructing measures. An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wright, B. D. (1999). Model selection: Rating Scale Model (RSM) or Partial Credit Model (PCM)? *Rasch Measurement Transactions*, 12(3), 641-642.
- Wright, B. D., y Linacre, J. M. (1989). Differences between scores and measures. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1, 1995* (pp. 63-65). Chicago, IL: MESA Press.
- Wright, B. D., y Linacre, J. M. (1994). Reasonable mean-square fit values. En J. M. Linacre (Ed.), *Rasch Measurement Transactions*, 1994 (p. 370). Chicago, IL: MESA Press.
- Wright, B. D., y Stone, M. (1998). *Diseño de mejores pruebas*. México: CENEVAL.
- Wright, B. D., y Masters, G. (1982). *Rating scale analysis*. Chicago, IL: MESA.
- 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>

Ministerio de Educación

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

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



PERÚ

Ministerio
de Educación