

SERIE
INVESTIGACIONES

LECTURA
MATEMÁTICA
APRENDIZAJES
COHORTE
PRIMARIA
FACTORES
ASOCIADOS
MODELO
LONGITUDINAL
EDUCACIÓN

MARCO DE FUNDAMENTACIÓN DE LAS PRUEBAS DE RENDIMIENTO

de la Evaluación Censal de Estudiantes
de 2.° de Secundaria 2015



PERÚ

Ministerio
de Educación



PROGRESO
PARA TODOS

LECTURA
MATEMÁTICA
APRENDIZAJES
COHORTE
PRIMARIA
FACTORES
ASOCIADOS
MODELO
LONGITUDINAL
EDUCACIÓN

MARCO DE FUNDAMENTACIÓN DE LAS PRUEBAS DE RENDIMIENTO

de la Evaluación Censal de Estudiantes
de 2.° de Secundaria 2015



PERÚ

Ministerio
de Educación



PERÚ

Ministerio
de Educación

MARCO DE FUNDAMENTACIÓN DE LAS PRUEBAS DE RENDIMIENTO DE LA EVALUACIÓN CENSAL DE ESTUDIANTES DE 2.º DE SECUNDARIA 2015

Responsable de la elaboración del documento

Humberto Hildebrando Pérez León Ibáñez

Participaron en la redacción del documento

Humberto Hildebrando Pérez León Ibáñez

Yoni Cristian Arámbulo Mogollón

Frank Joselín Villegas Regalado

Tulio Antonio Ozejo Valencia

Olimpia Rosa Castro Mora

Participaron en la revisión del documento

Tania Magaly Pacheco Valenzuela

Fernando José Llanos Masciotti

Andrés Alberto Burga León

María Elena Marcos Nicho

Correctora de estilo: Lérica Fernández

Diagramador: Lucía Escobedo

© Ministerio de Educación del Perú, 2015

Calle del Comercio 193, San Borja

Lima, Perú

Teléfono: 615-5800

www.minedu.gob.pe

Se autoriza citar o reproducir la totalidad o parte del presente documento, siempre y cuando se mencione la fuente.

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	8
1 Antecedentes	10
2 Marco normativo	11
3 Objetivos de la ECE	12
4 Principios que guían el diseño y construcción de los instrumentos de evaluación en la ECE	15
4.1 Validez y confiabilidad	15
4.1.1 Validez	15
4.1.2 Confiabilidad	16
4.2 Diseño Universal de Evaluación	17
5 Modelo general de la evaluación	19
6 Las pruebas de Comunicación	21
6.1 Justificación	21
6.2 Subprueba de Lectura	22
6.2.1 Definición de la competencia lectora	22
6.2.2 Modelo de evaluación: capacidades lectoras, textos y contextos	22
6.2.3 Diseño de la prueba	25
6.3 Subprueba de Escritura	26
6.3.1 Definición de la competencia en escritura	26
6.3.2 Modelo de evaluación: capacidades, textos y situaciones	27
6.3.3 Diseño de la prueba	28
7 La prueba de Matemática	31
7.1 Justificación	31
7.2 Definición de la competencia matemática	32
7.3 Modelo de evaluación: capacidades, contenidos y contextos	32
7.4 Diseño de la prueba	35

8 Características de la construcción de las pruebas	38
8.1 Tipos de ítems	38
8.2 Armado de las pruebas	40
9 Modelo de medición	42
9.1 El modelo Rasch	42
9.2 Ajuste al modelo	45
9.3 Funcionamiento diferencial de los ítems	45
Bibliografía	47

Índice de tablas

Tabla 4.1. Elementos del Diseño Universal de Evaluación	18
Tabla 6.1. Lectura. Distribución de ítems de la prueba por capacidad	25
Tabla 6.2. Lectura. Distribución de ítems de la prueba por tipo de texto y formato textual	26
Tabla 6.3. Escritura. Capacidades e indicadores evaluados en la primera consigna: anécdota (narrativo)	29
Tabla 6.4. Escritura. Capacidades e indicadores evaluados en la segunda consigna: artículo (descriptivo)	29
Tabla 6.5. Escritura. Capacidades e indicadores evaluados en la tercera consigna: noticia (narrativo)	30
Tabla 6.6. Escritura. Capacidades e indicadores evaluados en la cuarta consigna: carta (argumentativo)	30
Tabla 7.1. Matemática. Distribución de ítems de la prueba por capacidad a evaluar	35
Tabla 7.2. Matemática. Distribución de ítems de la prueba por contenido	36
Tabla 7.3. Matemática. Distribución de ítems de la prueba por contexto	37

Índice de gráficos

Gráfico 9.1. Curva característica para un ítem dicotómico	43
Gráfico 9.2. Curva característica para un ítem con un crédito parcial	44

Introducción

En la búsqueda de brindar un servicio educativo de calidad para todos, los sistemas educativos nacionales prestan atención a un conjunto amplio de indicadores: equidad en el acceso, horas lectivas efectivas, repitencia y deserción, etc. Entre ellos está el logro de los aprendizajes previstos por el currículo nacional, que es el documento que condensa lo que la sociedad espera para los niños y jóvenes en un recorrido que abarca once años de escolaridad.

Los aprendizajes de los estudiantes son monitoreados constantemente en las aulas por los docentes a través de una serie de técnicas e instrumentos de evaluación; sin embargo, para construir un indicador de alcance nacional que se aproxime a los aprendizajes mínimos esperados para todos, los sistemas educativos diseñan y aplican evaluaciones estandarizadas. Esta es una tarea que, por su envergadura, es de alta responsabilidad y complejidad. Es de alta responsabilidad porque sus resultados tienen repercusiones en el diseño de políticas educativas en varios niveles. Ello implica que la evaluación estandarizada debe constituirse en un proceso legitimado socialmente. En ese sentido, es también una tarea compleja, porque los mecanismos para conseguir dicha legitimidad (validez, confiabilidad, cobertura, estandarización, accesibilidad, probidad y confidencialidad) también lo son, tanto por los procesos altamente técnicos que se requieren como por la magnitud del operativo.

La evaluación, pues, requiere de un cuidadoso proceso de definición y planificación, de manera que los resultados reflejen con la mayor fidelidad posible las competencias reales de los estudiantes. El presente marco de trabajo constituye el primer escalón en este proceso de construcción de una evaluación de sistema, ya que desarrolla los principios, las definiciones fundamentales y las características de los instrumentos, los cuales guiarán las decisiones posteriores.

El primer capítulo es un resumen de las experiencias previas en evaluaciones censales, que permitirá poner en contexto esta evaluación.

El segundo capítulo menciona la legislación y normativa institucional en que se enmarca la evaluación.

El tercer capítulo desarrolla cada uno de los objetivos de la ECE y expone brevemente las razones de las principales decisiones tomadas.

El cuarto capítulo hace referencia a los principios que guían la construcción de las pruebas de la ECE, esto es, la validez, la confiabilidad y la accesibilidad.

El capítulo quinto explica el modelo general de evaluación, es decir, la relación entre contenidos, procesos y contextos que intervienen en la resolución de los ítems.

Los capítulos sexto y séptimo describen las características de las pruebas de Comunicación y Matemática, respectivamente, e incluyen las definiciones conceptuales y operacionales de los constructos evaluados y las proporciones de los ítems según los aspectos contemplados en el modelo de evaluación.

El octavo capítulo describe las características de la construcción de las pruebas, que incluyen los formatos de los ítems utilizados y la forma como se armaron los cuadernillos.

El noveno y último capítulo explica el modelo Rasch, el cual se emplea en la ECE para procesar las respuestas de los estudiantes. Este capítulo también incluye la descripción del armado de las pruebas.

Se espera que este nuevo esfuerzo de evaluación contribuya a una mejor comprensión de la realidad educativa y sirva para tomar cada vez más acertadas e informadas decisiones en este reto que significa mejorar los aprendizajes en nuestro país.

Antecedentes

La Oficina de Medición de la Calidad de los Aprendizajes del Ministerio de Educación (UMC) es la instancia responsable de diseñar e implementar las evaluaciones de logros de aprendizaje de los estudiantes de educación básica a nivel nacional. En esta lógica, conduce evaluaciones nacionales de carácter censal, así como coordina la aplicación de evaluaciones internacionales de índole muestral. Cada una de estas pruebas aporta con información relevante y complementaria sobre las distintas competencias e hitos del ciclo escolar.

En lo que respecta a las evaluaciones censales, se ha aplicado anualmente la ECE desde el 2007 (Ministerio de Educación - Oficina de Medición de la Calidad de los Aprendizajes, 2009) en dos grados: a) en 2.º grado de primaria, en las escuelas que atienden a poblaciones castellanohablantes; y b) en 4.º grado de primaria, en las escuelas que atienden a poblaciones hablantes de alguna lengua originaria y desarrollan el Programa de Educación Intercultural Bilingüe (EIB).

A partir del 2015, la ECE se aplicará por primera vez en 2.º de secundaria y medirá las competencias de Lectura, Escritura y Matemática. En los siguientes años, se incluirá la evaluación de otras competencias.

Marco normativo

La ECE es una evaluación de sistema que reporta oficialmente los resultados de aprendizaje de los estudiantes. Su existencia se fundamenta en diversos documentos legales nacionales y sectoriales que buscan asegurar el derecho de todos los estudiantes a una educación de calidad. A continuación listamos la normativa que sirve de base a la ECE:

- Ley General de Educación. Ley n.º 28044.
- Proyecto Educativo Nacional al 2021, aprobado por Resolución Suprema n.º 001-2007-ED.
- Diseño Curricular Nacional, aprobado por Resolución Ministerial n.º 0667-2005-ED.
- Resolución Ministerial n.º 199-2015-MINEDU, que modifica parcialmente el Diseño Curricular Nacional.
- Decreto Supremo n.º 021-2007-ED, que declara de preferente interés sectorial para el Ministerio de Educación la permanente implementación y ejecución de la evaluación del sistema educativo.
- Decreto Supremo n.º 015-2008-ED, que autoriza al Ministerio de Educación para disponer la implementación y ejecución de las evaluaciones del sistema educativo.
- Resolución Ministerial n.º 0554-2013-ED, que aprueba la Norma Técnica denominada "Norma para la implementación y ejecución a nivel nacional de Evaluaciones del Logro del Aprendizaje de los Estudiantes de Educación Básica".
- Decreto Supremo que modifica los D. S. n.ºs 021-2007 y 015-2008, que establecen la implementación y ejecución de las evaluaciones del sistema educativo.

Objetivos de la ECE

Objetivo 1. Dar información sobre los logros de aprendizaje en Comunicación (Lectura y Escritura) y Matemática de los estudiantes peruanos que cursan el 2.º grado de secundaria de la Educación Básica Regular.

La primera característica de esta evaluación es su carácter censal; es decir, su alcance abarca a todos los sujetos de una población definida. Esta evaluación ha sido concebida así porque se pretende devolver resultados a todos los actores de las instancias de gestión educativa: directores y docentes de las escuelas, especialistas y autoridades de las Unidades de Gestión Educativa Local (UGEL), Direcciones Regionales de Educación (DRE) y Ministerio de Educación (Minedu), con el fin de movilizarlos hacia la mejora de los aprendizajes en Lectura y Matemática.

Por su parte, la prueba de Escritura, que está incluida bajo el rótulo de la ECE, tiene carácter muestral, porque la calificación de las respuestas —a partir de redacciones largas que requieren correctores humanos que utilizan extensas rúbricas de evaluación— implica un proceso difícilmente aplicable —en las actuales condiciones— a toda la población. Por tanto, los resultados de la prueba de Escritura serán devueltos por estratos. Esta información complementará la que recibirán docentes, directores y demás actores educativos sobre Lectura y Matemática.

En su primera edición, la ECE de 2.º de secundaria evaluará Comunicación y Matemática, dado que son competencias fundamentales para el desenvolvimiento en la sociedad, así como necesarias para el desarrollo de otras competencias. La competencia comunicativa está presente en todas las situaciones de aprendizaje, sean estas escolares o extraescolares; por tanto, su dominio está a la base de las demás competencias. Por su parte, la competencia matemática —en el trabajo con cantidades, regularidades y relaciones espaciales— permite al estudiante resolver problemas reales, modelar su entorno y utilizar un lenguaje preciso para comunicar y argumentar ideas de manera crítica. Estas capacidades son básicas para la construcción de otros conocimientos, como el científico y el tecnológico, cuyas bases también se enseñan en la escuela. Debido a que esta evaluación debe ofrecer

un panorama que sea lo más completo posible del estado de los aprendizajes, se tiene como meta incorporar en las siguientes ediciones evaluaciones de las competencias vinculadas a las ciencias sociales y las ciencias naturales.

El grado elegido, 2.º de secundaria, marca el final del VI ciclo de la Educación Básica Regular y se ubica en un punto intermedio de la educación secundaria. Habiendo consolidado los aprendizajes propios del nivel primario y los básicos del presente nivel, los estudiantes ingresan a la fase final de su recorrido escolar, que exigirá de ellos el desarrollo de pensamientos más complejos y abstractos. La información recogida en este punto resulta estratégica para que los distintos actores del sistema educativo realicen ajustes oportunos en estrategias, metodologías e intervenciones pedagógicas e institucionales, de manera que los estudiantes culminen exitosamente su escolaridad obligatoria.

Objetivo 2. Comparar en el tiempo los resultados, para informar sobre la evolución de los aprendizajes.

La ECE de 2.º de secundaria está proyectada para producir información anual sobre el rendimiento escolar. La periodicidad y las características del diseño permitirán trazar tendencias a lo largo del tiempo para saber si los resultados están mejorando o si las brechas entre las distintas poblaciones de estudiantes se están estrechando.

Objetivo 3. Generar evidencia sobre las características individuales y familiares de los estudiantes evaluados, así como de las condiciones y los procesos escolares que se encuentran asociados a su desempeño.

El rendimiento escolar reportado por la evaluación es el producto de las múltiples interrelaciones entre los actores educativos, las instituciones y los procesos históricos y culturales, entre otros aspectos. Para el sistema educativo, entendido en un sentido amplio, es importante aproximarse al conocimiento de estos fenómenos, esto es, a las posibles relaciones que explican las variaciones en el rendimiento. Esta información permite profundizar la indagación en líneas de investigación específicas o la toma de decisiones de política sobre la base de evidencia empírica.

Para ello, en la ECE, además de las pruebas de rendimiento, se aplicará un conjunto de cuestionarios para recoger información sobre el nivel socioeconómico, la trayectoria del estudiante, el clima escolar y las prácticas pedagógicas en el aula. A escala muestral, junto con las pruebas de Escritura, se recogerá información sobre el liderazgo pedagógico del director

y diversas concepciones del docente acerca de temas de su especialidad (Comunicación o Matemática, según sea el caso).

El presente marco de fundamentación aborda lo concerniente al primer y segundo objetivo. Las particularidades de los instrumentos y las metodologías para recoger y procesar información sobre los factores asociados se detallarán en otro documento.

Principios que guían el diseño y construcción de los instrumentos de evaluación en la ECE

El diseño y la construcción de instrumentos son aspectos cruciales en la evaluación. Ellos deben recoger información de los estudiantes de tal manera que se pueda estimar de forma fiable su nivel de aprendizaje y que esa información pueda ser usada para los fines propios del proceso educativo. Asimismo, debe reflejar una concepción inclusiva de la educación, conforme a los lineamientos de la política educativa nacional. Ello implica que el diseño debe buscar eliminar las barreras que impidan a los estudiantes demostrar sus reales competencias. En este capítulo, se abordarán tres grandes principios que guían la construcción de los instrumentos de evaluación en la ECE: validez, confiabilidad y diseño universal de evaluación.

4.1 Validez y confiabilidad

4.1.1 Validez

La validez es un concepto unitario que alude al grado en que la evidencia empírica y la teoría apoyan la interpretación y uso de las puntuaciones que son resultado de la aplicación de una prueba. Así entendida, esta definición se aleja de las aproximaciones tradicionales por las cuales la validez era una propiedad del instrumento (AERA, APA y NCME, 2014; Messick 1989, 1990, 1995).

El concepto de validez implica que se deben tener intenciones claras y declaradas sobre dichas interpretaciones y consecuencias, de manera que estas puedan ser contrastadas con la evidencia disponible. A su vez, supone definir con precisión el o los constructos a evaluar (AERA *et al.*, 2014). Aproximaciones recientes señalan que la delimitación del constructo debe responder a un modelo de aprendizaje; esto es, debe quedar establecida la manera como este progresa. La claridad sobre este progreso orientará la posterior elaboración de tareas cognitivas cuyas respuestas serán usadas en la interpretación de resultados (National Research Council, 2001). En el caso de la ECE, los constructos y la lógica de su progresión están bien asentados en la literatura existente, en los documentos curriculares (Diseño Curricular Nacional, R. M. n.º 199-2015-MINEDU, Mapas de Progreso del Aprendizaje) y en otros documentos de uso pedagógico, como las Rutas del Aprendizaje.

A lo largo del proceso del diseño y construcción de instrumentos de la ECE, se establecen múltiples puntos en los cuales se obtienen evidencias de validez que están referidas a distintos aspectos. Por ejemplo, se recogen evidencias de validez vinculadas al contenido cuando las tablas de especificaciones y los ítems son sometidos a juicio de expertos para determinar si los indicadores y los ítems conforman una muestra adecuada del constructo definido. Asimismo, la construcción de los ítems es puesta a prueba tanto en procesos de juicio experto como en aplicaciones de campo, para verificar, en el primer caso, que a) las tareas propuestas se ajusten a los indicadores, y que b) se minimice el sesgo o las fuentes de varianza irrelevante. La aplicación de campo permite obtener evidencias sobre la validez de la estructura unidimensional del constructo a partir de los patrones de las respuestas a la prueba, así como posibles sesgos en la construcción de los ítems, evidenciados por su comportamiento diferencial. Los detalles de los procedimientos psicométricos y estadísticos con los cuales se obtiene esta evidencia se encuentran en el capítulo 9.

4.1.2 Confiabilidad

La confiabilidad se entiende como la consistencia de los resultados de una prueba a través de múltiples aplicaciones independientes (AERA *et al.*, 2014; Brennan, 2001; Carmines y Zeller, 1979). Este concepto está ligado al de validez, en la medida en que la interpretación de los rangos aceptables de consistencia tiene que ver con la naturaleza de las interpretaciones y usos que se desea dar a los puntajes de la prueba (AERA *et al.*, 2014). Aunque existen varios procedimientos para estimar la confiabilidad directamente (por ejemplo, *test-retest* o formas paralelas), en la práctica, particularmente en evaluaciones estandarizadas masivas, estos métodos no siempre están disponibles. Por ello, se utilizan métodos indirectos que calculan la consistencia interna sobre la base del promedio de las correlaciones entre los ítems, o como una derivación del cociente entre la sumatoria de varianzas de los ítems sobre la varianza total (en teoría clásica) o de la separación de las personas (en el caso de los modelos Rasch) (Linacre, 2014). Este último criterio es el que utiliza la ECE para estimar la confiabilidad de las puntuaciones derivadas de un instrumento de evaluación. Adicionalmente, tanto en la codificación de preguntas de respuesta construida en Lectura y Matemática como en las pruebas de Escritura, se calculan coeficientes de consistencia interevaluadores en procedimientos de corrección múltiple, doble corrección y precisión.

4.2 Diseño Universal de Evaluación

El Diseño Universal de Evaluación es una extensión de los principios del Diseño Universal usado inicialmente en la arquitectura al ámbito de la educación y, en particular, a la evaluación a gran escala (Hanna, 2005). El Diseño Universal es a la vez un movimiento y un conjunto de lineamientos a favor del acceso de todos, incluyendo a las personas con discapacidad, a las características de las construcciones (Story, Mueller, y Mace 2011). Los principios del Diseño Universal incluyen lo siguiente: uso equitativo para personas con diversas habilidades; uso flexible e intuitivo, información necesaria para el uso, tolerancia al error, mínimo esfuerzo físico, y tamaño y espacio favorable al uso y la manipulación (Story *et al.*, 2011).

En el ámbito educativo, se ha desarrollado el Diseño Universal para el Aprendizaje, el cual parte por reconocer que en la enseñanza tradicional muchos estudiantes, no solo aquellos con discapacidad, enfrentan barreras que les impiden desarrollar sus habilidades de manera óptima (Meyer, Rose, y Gordon 2013). El Diseño Universal para el Aprendizaje tiene por objetivo la flexibilización de las estrategias de enseñanza, de tal manera que se puedan atender las múltiples necesidades educativas de los estudiantes (Hanna, 2005). Para ello propone básicamente tres lineamientos, que implican ofrecer múltiples medios para a) promover el interés, b) promover el conocimiento y c) promover la acción y la expresión (National Center On Universal Design for Learning, n. d.).

En particular, el Diseño Universal de Evaluación tiene por finalidad establecer guías para que el diseño de las pruebas pueda incluir al más amplio rango de estudiantes, y que se puedan hacer inferencias válidas sobre su desempeño (Thompson, Johnstone y Thurlow, 2002; Thompson y Thurlow, 2002).

Tabla 4.1. Elementos del Diseño Universal de Evaluación

Elemento	Explicación
Evaluación inclusiva de la población	El diseño de pruebas debe pensarse para incluir a todos los estudiantes.
Constructos definidos con precisión	Los constructos específicos deben ser claramente definidos de tal manera que sean removidas todas las barreras cognitivas, sensoriales, emocionales y físicas que son irrelevantes para el constructo.
Ítems accesibles, no sesgados	La accesibilidad debe ser concebida desde el inicio del diseño. Los procedimientos de revisión de sesgos deben asegurar la calidad de todos los ítems.
Susceptibles de acomodaciones	El diseño de los ítems debe facilitar el uso de acomodaciones (por ejemplo, debe facilitar su transcripción al sistema Braille).
Instrucciones y procedimientos simples, claros e intuitivos	Todas las instrucciones deben presentarse de forma clara y entendible.
Máxima legibilidad y comprensibilidad	Se debe procurar que el lenguaje sea llano, que la longitud de las oraciones sea razonable. Asimismo, se debe procurar que las tablas, textos, figuras e ilustraciones aseguren su descifrabilidad.

Fuente: Thompson y Thurlow (2002).

Con el fin de que la ECE sea una evaluación inclusiva, la UMC ha adoptado los principios del Diseño Universal de Evaluación en la construcción de sus instrumentos. Progresivamente, se irán implementando acomodaciones, de modo que estudiantes con necesidades educativas especiales (NEE) puedan también ser incluidos en la evaluación.

Modelo general de la evaluación

Toda situación de evaluación requiere que se recoja información suficiente, que permita una aproximación adecuada al constructo que se desea evaluar. Para ello, es necesario concebir un modelo de cómo se organiza conceptualmente la variable de interés, para identificar aquellos aspectos que son centrales en la definición de la variable.

El aprendizaje es un fenómeno complejo, pues involucra no solo la esfera cognitivo-racional, sino también la social y la afectiva. La evaluación de logros de aprendizaje debe aproximarse a esta complejidad desde la concepción de un modelo que la refleje de alguna manera.

El enfoque de evaluación de cada competencia de la ECE se organiza en tres dimensiones que se explican a continuación:

- **Capacidades.** Son los procesos cognitivos que el estudiante pone en juego para realizar eficazmente tareas que pretenden simular situaciones de la vida cotidiana, tales como representar, razonar, deducir, analizar o resolver problemas. En el caso particular de las pruebas, las capacidades evaluadas son aquellas que el estudiante requiere para enfrentar exitosamente las preguntas propuestas.
- **Contenidos.** Son los conocimientos disciplinares vinculados a las competencias curriculares o, en el caso de la lectura, los soportes textuales que necesita el estudiante para desplegar o poner en juego determinados procesos. Dichos contenidos a evaluar han sido tomados de los documentos curriculares vigentes al momento de diseñar las pruebas.
- **Contextos.** Son las diversas situaciones o escenarios en los que el estudiante pone en práctica las capacidades articuladas con los contenidos al momento de ser evaluado. Estos contextos pueden ser propios de la vida escolar, personal, pública, etc.

La consideración de estos tres elementos en el diseño de la prueba provee una amplia variedad de tareas y formatos, desde los más simples hasta los más complejos, lo cual permite contar con mayores y mejores fuentes de validez de contenido y constructo.

Aunque se sabe que el componente afectivo es indesligable de los procesos de aprendizaje, el modelo de evaluación de la ECE no lo incorpora todavía y no será recogido en las pruebas. Se está estudiando la posibilidad, en el futuro, de explorar este componente con mayor detalle. Sin embargo, para esta evaluación, se recogerá información sobre el clima escolar, una variable que involucra actitudes y que en Latinoamérica ha demostrado tener una considerable potencia explicativa de los resultados de aprendizaje (Treviño, Valdés, Castro y Costilla, 2010).

En los capítulos siguientes, se describirán las características de las pruebas que conforman la ECE de 2.º de secundaria. Como ya se señaló, esta construcción se ajusta a la normativa curricular vigente: Diseño Curricular Nacional y R. M. n.º 199-2015-MINEDU.

Las pruebas de Comunicación

En este capítulo, se desarrollarán las características de las pruebas de Comunicación. En la ECE, las competencias de Comunicación que se evalúan son Lectura y Escritura. Como se indicó, la prueba de Escritura, por la complejidad de su procesamiento, se aplica en una muestra representativa nacional de estudiantes. Ello implica que los resultados no serán reportados a nivel de institución educativa (IE), sino de estratos nacionales.

6.1 Justificación

El mundo contemporáneo es mayoritariamente un mundo letrado. La lectura y escritura son productos culturales y sociales que, desde su invención en adelante, han ido mediando progresivamente la comunicación entre las personas de la mayoría de los pueblos y que, en la actualidad, se han hecho imprescindibles para la vida ciudadana: la educación, el trabajo, la economía, el desarrollo de la ciencia y la tecnología, y en general, el ejercicio de los derechos.

Más aún, la producción de información ha generado nuevas formas de usar la lectura y la escritura. Por ejemplo, el manejo de múltiples fuentes, datos organizados en tablas y gráficos, información espacialmente organizada y la proliferación de medios digitales han configurado nuevos retos para la alfabetización de las personas.

En ese sentido, el aprendizaje de la lectura y la escritura es un imperativo para que las personas puedan desenvolverse con suficiencia en la sociedad y ejercer plenamente su ciudadanía. La escuela, en este escenario, es una institución clave para desarrollar en los estudiantes capacidades que les permitan insertarse con éxito en la sociedad y contribuir con ella.

En particular, la educación secundaria es un espacio idóneo para el perfeccionamiento de las capacidades comunicativas. En los estudiantes adolescentes, las características del desarrollo cognitivo y social hacen posible la consolidación de las capacidades lectoras y de escritura que fueron adquiridas durante la primaria, a través de un largo proceso, y se observan aptitudes para desarrollar capacidades de procesamiento y uso

de información más complejas, es decir, más abstractas, menos apegadas a lo inmediato y que involucran más fuentes.

En el VI ciclo de la Educación Básica Regular (EBR), la escuela debe promover con creciente énfasis situaciones de lectura que asemejen las complejidades de la realidad circundante, como la confrontación de múltiples fuentes y el análisis intertextual (Britt, Goldman y Rouet, 2013). Asimismo, debe desarrollar el distanciamiento y la toma de postura con respecto del texto (Cassany y Castellà, 2010; Cassany, 2003), de manera que el lector pueda construir su opinión con autonomía y suficiencia argumental. En el caso de la escritura, la escuela debe promover el manejo de diversos tipos y géneros como herramientas de comunicación efectiva para resolver problemas comunicativos reales.

6.2 Subprueba de Lectura

En este apartado se describen las características de la prueba de Lectura.

6.2.1 Definición de la competencia lectora

En la ECE se concibe la lectura como un proceso dinámico de construcción de significados (Kintsch y Mangalath, 2011) que implica el despliegue de un conjunto de habilidades que permiten al lector relacionar e integrar información con la finalidad de comprender el texto en su totalidad, así como reflexionar sobre el texto leído y emitir un juicio crítico sobre algunos de sus elementos. Dicho proceso ocurre en un contexto social, en el cual los textos son mediadores y, a la vez, producto de las prácticas letradas de una comunidad (Barton y Hamilton, 2000).

La lectura es un proceso estratégico e interactivo. Es un proceso estratégico porque el lector decide rutas, atajos y procedimientos para comprender el sentido del texto de acuerdo a sus propósitos de lectura (Cubukcu, 2007; Garner, 1987). Es un proceso interactivo porque el lector elabora continuamente hipótesis y deducciones que va contrastando con la información del texto y de su saber previo sobre el mundo, y a su vez utiliza la información como insumo para elaborar nuevas hipótesis (Perfetti y Roth, 1981; Rumelhart y McClelland, 1981).

6.2.2 Modelo de evaluación: capacidades lectoras, textos y contextos

De acuerdo con el modelo de evaluación adoptado, en la ECE la lectura se evalúa a partir de tres dimensiones: capacidades lectoras, textos y contextos de lectura.

a. Capacidades lectoras

Las capacidades lectoras son las habilidades que el lector despliega cuando lee un texto. Las tareas propuestas en la evaluación buscan que el estudiante reproduzca las capacidades que utilizaría en situaciones reales en las que debe conseguir ciertas metas de comprensión para resolver un problema comunicativo.

Las capacidades lectoras consideradas en la ECE están en correspondencia con la R. M. n.º 199-2015 y definidas según las Rutas del Aprendizaje (Ministerio de Educación del Perú, 2015a) de la siguiente manera:

- *Recupera información explícita del texto.* Consiste en localizar “información que se presenta en el texto de manera explícita sin necesidad de hacer inferencias. A partir de esta, discrimina la que requiere según su interés y propósito” (p. 113).
- *Infiere el significado del texto.* Consiste en construir significados o relaciones no explícitas en el texto, pero necesarias para su comprensión. El estudiante formula estas inferencias “a partir de sus saberes previos, de los indicios que le ofrece el texto y del contexto en el que este se produce. Mientras va leyendo, verifica o reformula sus hipótesis de lectura” (p. 116).
- *Reflexiona sobre la forma, el contenido y el contexto del texto.* Mediante esta capacidad, el estudiante “toma distancia de las ideas propuestas en el texto, o de los recursos utilizados para transmitir ese significado, y juzga si son adecuados o no. Para ello, considera objetivamente el contenido y la forma, evalúa su calidad y adecuación con una perspectiva crítica. Opina reflexivamente sobre el texto usando argumentos que demuestren si lo comprendió” (p. 119).

b. Textos

En la ECE, los textos son clasificados a partir de dos criterios: tipos textuales y formatos textuales.

Tipos textuales. Los textos se componen de secuencias proposicionales, las cuales, al encadenarse, van construyendo el discurso. Una secuencia textual es una secuencia de proposiciones que, en conjunto, tienen un propósito retórico específico: narrar, convencer, describir, instruir, etc. Puede haber varias secuencias textuales en un mismo texto

(Adam, 2001). En un texto completo, el tipo textual corresponde con la secuencia textual dominante.

En la ECE, los tipos textuales considerados, de acuerdo con la clasificación de Werlich (1976), son:

- *Narrativos*. Organizan la información como una secuencia de hechos sucesivos que ocurren en un espacio y un tiempo.
- *Expositivos*. Proporcionan una explicación causal sobre el modo en que distintos elementos de un sistema se relacionan.
- *Descriptivos*. Ofrecen información sobre cómo es o ha sido una persona, animal, objeto o espacio.
- *Argumentativos*. Demuestran al lector la validez de una idea o punto de vista a partir del desarrollo de razones o argumentos que posibilitan la persuasión.
- *Instructivos*. Presentan un conjunto de órdenes o recomendaciones para la realización de una actividad específica.

Formatos textuales. El formato es la forma concreta como se organizan las ideas en un soporte físico. La disposición de las frases, oraciones, datos, imágenes y otros elementos colabora en la conformación del mensaje. Los formatos considerados en la ECE son:

- *Continuo*. Es aquel texto que se compone normalmente de una serie de oraciones organizadas en párrafos.
- *Discontinuo*. Se organiza de manera diferente al texto continuo (en columnas, tablas, gráficos, etc.) y, por lo tanto, requiere otro modo de leer.
- *Mixto*. Presenta algunas secciones continuas y otras discontinuas, sin perder su carácter unitario.
- *Múltiple*. Incluye dos textos provenientes de dos fuentes o autores diferentes que, para los fines de la evaluación, son dispuestos como parte de una misma situación comunicativa.

c. Contextos

Los contextos son las situaciones sociales en las que es usada la lectura. Son importantes porque ponen en evidencia el sentido de la lectura como práctica social. En la evaluación se incluyen textos que

típicamente pueden ser encontrados en diversos contextos en el mundo real. Los contextos considerados en la ECE son:

- *Educacional*. Se corresponde con situaciones de instrucción formal o informal, con el propósito de generar aprendizaje.
- *Público*. El texto es accesible para el público en general y es de interés general.
- *Recreacional*. El texto es usado con objetivos de esparcimiento y disfrute.

Cabe señalar que esta clasificación según contextos no es rígida (podrían proponerse otras situaciones de lectura) y, por otra parte, que un mismo texto puede obedecer a uno o más contextos de lectura. En esta evaluación, la categorización de contextos es utilizada referencialmente para conseguir mayor diversidad en la selección de textos.

6.2.3 Diseño de la prueba

En esta sección, se muestra la distribución de ítems por cada aspecto del modelo de evaluación en la prueba de Lectura.

a. Capacidades

La Tabla 6.1 muestra la distribución de preguntas por capacidad en la prueba de Lectura.

Tabla 6.1. Lectura. Distribución de ítems de la prueba por capacidad

Capacidad	Cantidad de ítems	
	n	%
Recupera información explícita del texto.	14	18,4
Infiere el significado del texto.	43	56,6
Reflexiona sobre la forma, el contenido y el contexto del texto.	19	25,0
TOTAL	76	100,0

En la prueba de lectura de la ECE se da prioridad a la capacidad de inferir por sobre la de recuperar información explícita, dado que este proceso juega un papel fundamental en la comprensión del discurso y es esperable que esté más desarrollado que al finalizar la primaria. La mera comprensión literal no garantiza una plena construcción del significado del texto, sino que esta depende de realizar múltiples inferencias que conectan los significados de los enunciados que se tejen al interior del texto (microestructura) y a nivel global del texto (macroestructura) con el conocimiento del lector sobre el mundo y el lenguaje mismo. Se trata de una capacidad clave que permite acceder a distintos niveles de representación mental del discurso.

b. Textos

La Tabla 6.2 muestra la distribución de preguntas por tipo y formato textual en la prueba de Lectura.

Tabla 6.2. Lectura. Distribución de ítems de la prueba por tipo de texto y formato textual

Tipos textuales	Formatos textuales							
	Continuo		Discontinuo		Mixto		Múltiple	
	n	%	n	%	n	%	n	%
Narrativos	5	14,7	6	28,6	0	0,0	5	50,0
Expositivos	6	17,7	10	47,6	6	54,6	0	0,0
Descriptivos	7	20,6	0	0,0	5	45,5	0	0,0
Argumentativos	11	32,3	0	0,0	0	0,0	5	50,0
Instructivos	5	14,7	5	23,8	0	0,0	0	0,0
TOTAL	34	100,0	21	100,0	11	100,0	10	100,0

6.3 Subprueba de Escritura

6.3.1 Definición de la competencia en escritura

La escritura es la habilidad de un emisor para organizar en un texto escrito ideas que responden a una meta comunicativa. El acto de escribir es tanto un proceso cognitivo como social. Es cognitivo en tanto se pone en juego una serie de operaciones mentales para interpretar el contexto, fijar metas, codificar información y monitorear dichos procesos (Deane

et al., 2008; Flower y Hayes, 2014; Hayes, 2000). Es social en tanto es un acto inmerso en una red de relaciones sociales y convenciones culturales que establecen los géneros, estructuras y acuerdos discursivos o pragmáticos que delimitan los parámetros de la competencia. En este contexto, la escritura cumple distintas funciones sociales, principalmente la de comunicar a otros, pero también cumple otras conexas como ayudar a organizar el pensamiento, capturar información oral y escrita para uso personal, etc. (Barton y Hamilton, 2000; Kucer, 2005).

6.3.2 Modelo de evaluación: capacidades, textos y situaciones

En la ECE, la escritura se evalúa a partir de tres aspectos: capacidades, textos y situaciones comunicativas.

a. Capacidades

Las capacidades refieren a las habilidades que el estudiante despliega al momento de escribir textos.

- *Adecúa el texto considerando los posibles lectores, el propósito y el tipo de texto.* Esta capacidad se manifiesta cuando el estudiante produce un texto adecuado al contexto de comunicación en relación con el tema, el tipo de texto y el registro.
- *Organiza las ideas de su texto de manera coherente.* Esta capacidad consiste en expresar las ideas de forma clara y consistente, de tal manera que el texto producido pueda ser entendido por el lector. Para que el texto sea coherente, debe presentar un tema central que se mantiene y se desarrolla sin digresiones, vacíos de información, repeticiones innecesarias de ideas o contradicciones.
- *Usa de manera pertinente distintos mecanismos de cohesión.* Esta capacidad consiste en enlazar las ideas para mostrar sus relaciones. Para ello, el estudiante se vale de recursos como conectores, referentes y signos de puntuación.
- *Desarrolla ideas e información variadas.* Esta capacidad consiste en la expansión de las ideas básicas de un texto a través de precisiones, aclaraciones o detalles que brindan mayor riqueza comunicativa.
- *Usa convenciones ortográficas.* Esta capacidad consiste en el uso correcto de convenciones ortográficas tales como la tildación y el uso de grafías.

b. Textos

En la ECE, se propone a los estudiantes de 2.º de secundaria producir cuatro tipos de texto: dos narrativos, uno descriptivo y uno argumentativo. A continuación, se describen las características de cada uno de los textos solicitados.

- El primer texto solicitado es una anécdota en la que el estudiante narra una secuencia de eventos sobre una situación particular (tipo: narrativo; género: anécdota).
- El segundo texto solicitado es un artículo en el que el estudiante describe un objeto, una persona o un lugar de su entorno (tipo: descriptivo; género: artículo).
- El tercer texto solicitado es una noticia en la que el estudiante narra un hecho ocurrido en su comunidad (tipo: narrativo; género: noticia).
- El cuarto texto solicitado es una carta en la que el estudiante toma una posición a favor o en contra de alguna afirmación o hecho y la sustenta mediante argumentos (tipo: argumentativo; género: carta).

c. Situaciones

La prueba de escritura plantea al estudiante de 2.º de secundaria cuatro consignas, es decir, tareas que describen una situación comunicativa que sirve de estímulo para la producción de un texto (Calfée y Miller, 2007). Adicionalmente, cada una de las consignas indica claramente el propósito, la audiencia (los posibles lectores del texto) y el tipo textual que se solicita producir al estudiante. Estas consignas han sido construidas considerando el grado y los intereses de los estudiantes según su edad, de tal manera que se sientan motivados a construir textos completos. Asimismo, se han construido consignas que no requieren conocimientos especializados y que son próximas a las situaciones del entorno de los estudiantes.

6.3.3 Diseño de la prueba

La competencia en escritura se evalúa mediante consignas, que son enunciados que, a partir de una situación simulada de la realidad, instan al estudiante a resolver un problema comunicativo mediante la redacción de un escrito.

Cada estudiante se enfrenta a cuatro consignas que están relacionadas temáticamente y que requieren estructurar los tipos de secuencias textuales comentadas en el apartado anterior. Asimismo, las consignas suponen distintos niveles de demanda cognitiva para los estudiantes, para que se recojan evidencias de su habilidad ante tareas con diferentes grados de dificultad.

Las producciones son codificadas mediante rúbricas analíticas, las cuales miden la calidad del texto a partir de diversos criterios. A continuación, se señala qué criterios son evaluados en las rúbricas por cada consigna:

Tabla 6.3. Escritura. Capacidades e indicadores evaluados en la primera consigna: anécdota (narrativo)

Capacidades	Indicadores evaluados
Adecúa su texto a la situación comunicativa.	- Adecuación al tema
Organiza las ideas de su texto de manera coherente.	- Suficiencia comunicativa
Usa de manera pertinente distintos mecanismos de cohesión.	- Cohesión - Puntuación
Usa convenciones ortográficas.	- Escritura de palabras - Tildación

Tabla 6.4. Escritura. Capacidades e indicadores evaluados en la segunda consigna: artículo (descriptivo)

Capacidades	Indicadores evaluados
Adecúa su texto a la situación comunicativa.	- Adecuación a la secuencia descriptiva y al tema
Organiza las ideas de su texto de manera coherente.	- Coherencia
Usa de manera pertinente distintos mecanismos de cohesión.	- Cohesión - Puntuación
Desarrolla ideas e información variadas.	- Desarrollo descriptivo
Usa convenciones ortográficas.	- Escritura de palabras - Tildación

Tabla 6.5. Escritura. Capacidades e indicadores evaluados en la tercera consigna: noticia (narrativo)

Capacidades	Indicadores evaluados
Adecúa su texto a la situación comunicativa.	- Adecuación a la secuencia narrativa y al tema - Registro - Género noticia
Organiza las ideas de su texto de manera coherente.	- Coherencia
Usa de manera pertinente distintos mecanismos de cohesión.	- Cohesión - Puntuación
Desarrolla ideas e información variadas.	- Desarrollo descriptivo
Usa convenciones ortográficas.	- Escritura de palabras - Tildación

Tabla 6.6. Escritura. Capacidades e indicadores evaluados en la cuarta consigna: carta (argumentativo)

Capacidades	Indicadores evaluados
Adecúa su texto a la situación comunicativa.	- Adecuación a la secuencia textual argumentativa - Adecuación al tema - Registro
Organiza las ideas de su texto de manera coherente.	- Suficiencia comunicativa - Coherencia
Usa de manera pertinente distintos mecanismos de cohesión.	- Cohesión - Puntuación
Desarrolla ideas e información variadas.	- Desarrollo argumentativo - Plausibilidad argumentativa - Contraargumentación
Usa convenciones ortográficas.	- Escritura de palabras - Tildación

Las pruebas de Matemática

En este capítulo describiremos las características de la prueba de Matemática en la ECE de 2.º de secundaria.

7.1 Justificación

La matemática ha servido y sirve a las personas y sociedades para modelar el mundo alrededor, solucionar problemas de subsistencia y desarrollar tecnología (Cooke, 2013; Merzbach y Boyer, 2011). En ese sentido, el conocimiento matemático ha sido clave para organizar la realidad en que se vive.

En el mundo contemporáneo, existe un acelerado desarrollo de la ciencia y la tecnología, y un incremento exponencial de los volúmenes de información numéricamente expresada (Quantitative Literacy Design Team, 2006). Esta realidad demanda una ciudadanía que piense, comprenda y se exprese con suficiencia en términos matemáticos, para que pueda ejercer adecuadamente sus derechos y contribuir al desarrollo (The National Council on Education and the Disciplines, 2006).

Si bien esta competencia se desarrolla en situaciones reales que requieren resolver problemas concretos, la experiencia cotidiana por sí misma no es suficiente para formar el pensamiento matemático necesario para enfrentar los retos de la sociedad contemporánea (Quantitative Literacy Design Team, 2006). La educación formal escolar, en ese sentido, debe brindar las oportunidades para el acceso a las competencias matemáticas que exigen tanto las situaciones cotidianas del entorno inmediato como aquellas más lejanas e hipotéticas, e incluso aquellas que son puramente formales y abstractas, necesarias para el aprendizaje continuo (National Council of Teachers of Mathematics, 2000a) en diversos contextos culturales.

La educación matemática de los estudiantes del VI ciclo de EBR debe posibilitarles que afiancen su autonomía (intelectual, afectiva y moral), valoren y desarrollen la riqueza expresiva del lenguaje matemático, y potencien el desarrollo del pensamiento de modo que este sea más flexible, estructurado y abstracto. Asimismo, debería propender a ampliar

su horizonte numérico (delimitando como nuevo universo los números racionales), desarrollar un uso extenso y aplicado del lenguaje algebraico, estimular un manejo comprensivo y aplicado de la diversidad de relaciones de las formas geométricas, y también fomentar un uso sistemático del razonamiento probabilístico en situaciones de incertidumbre.

7.2 Definición de la competencia matemática

La competencia matemática es un saber actuar deliberado y reflexivo que selecciona y moviliza una diversidad de habilidades, conocimientos matemáticos, destrezas, actitudes y emociones, en la formulación y resolución de problemas en una variedad de contextos.

Esta competencia se pone de manifiesto en situaciones referidas a cuantificar, medir, identificar regularidades, establecer equivalencias y variaciones, caracterizar y describir la forma y ubicación de los objetos; asimismo en la organización y sistematización de datos, el manejo de la incertidumbre, entre otros.

7.3 Modelo de evaluación: capacidades, contenidos y contextos

En la ECE, la competencia matemática se evalúa a partir de tres aspectos: capacidades, contenidos y contextos.

a. Capacidades

Las capacidades son las habilidades cognitivas que el estudiante tiene para adquirir y aplicar los conocimientos matemáticos (National Council of Teachers of Mathematics, 2000b).

Las capacidades consideradas en la ECE están en correspondencia con la R. M. n.º 199-2015 y toma las definiciones de las Rutas del Aprendizaje (Ministerio de Educación del Perú, 2015b):

- *Matematiza situaciones.* Es “expresar un problema, reconocido en una situación, en un modelo matemático” (p. 29). Esta capacidad implica el reconocimiento de características, datos, variables y condiciones del problema para dar curso a la construcción de un sistema que reproduzca o imite la realidad; asimismo, implica el uso del modelo en otras situaciones y la evaluación de sus alcances y limitaciones.
- *Comunica y representa ideas matemáticas.* Se trata de “comprender

el significado de las ideas matemáticas, y expresarlas de forma oral y escrita usando el lenguaje matemático y diversas formas de representación con material concreto, gráfico, tablas, símbolos y recursos TIC, y transitando de una representación a otra” (p. 30). Por ejemplo, la interpretación y el uso de gráficos, tablas o diagramas son parte de las habilidades que corresponden a esta capacidad.

- *Elabora y usa estrategias.* Se refiere a la capacidad de “planificar, ejecutar y valorar una secuencia organizada de estrategias y diversos recursos, entre ellos las tecnologías de información y comunicación, empleándolas de manera flexible y eficaz en el planteamiento y resolución de problemas, incluidos los matemáticos” (2015b, p. 32). Las habilidades que están involucradas en la capacidad son el diseño de un plan para solucionar el problema, la selección de procedimientos heurísticos, de cálculo mental, etc., y la valoración de las estrategias que fueron utilizadas.
- *Razona y argumenta generando ideas matemáticas.* Es “la capacidad de plantear supuestos, conjeturas e hipótesis de implicancia matemática mediante diversas formas de razonamiento (deductivo, inductivo y abductivo), así como el verificarlos y validarlos usando argumentos” (2015b, p. 33).

b. Contenidos

Los contenidos constituyen el cuerpo de conocimientos que sustenta la Matemática. Están indisolublemente relacionados con las capacidades. Al respecto, el NCTM señala que “no se pueden resolver problemas sin comprender y usar contenidos matemáticos. El conocimiento geométrico requiere razonamiento. Los conceptos algebraicos pueden analizarse y comunicarse por medio de representaciones” (National Council of Teachers of Mathematics, 2000a, pp. 7-8). En la ECE, se consideran cuatro organizadores de contenidos matemáticos, en correspondencia con la R. M. n.º 199-2015, definidos del siguiente modo según las Rutas del Aprendizaje (Ministerio de Educación del Perú, 2015b):

- *Cantidad.* Su desarrollo está referido a “modelos de solución numérica, comprendiendo el sentido numérico y de magnitud, la construcción del significado de las operaciones, así como la aplicación de diversas estrategias de cálculo y estimación al

resolver un problema” (p. 19).

- *Regularidad, equivalencia y cambio.* Comprende “desarrollar progresivamente la [...] patrones, la comprensión y el uso de igualdades y desigualdades, y la comprensión y el uso de relaciones y funciones. Toda esta comprensión se logra usando el lenguaje algebraico como una herramienta de modelación de distintas situaciones de la vida real” (p. 22).
- *Forma, movimiento y localización.* En este organizador, se considera el desarrollo del “sentido de la ubicación en el espacio, la interacción con los objetos, la comprensión de propiedades de las formas y cómo estas se interrelacionan, así como la aplicación de estos conocimientos al resolver diversos problemas” (2015b, p. 24).
- *Gestión de datos e incertidumbre.* Este organizador considera conocimientos relativos al procesamiento, representación e interpretación de datos, que son transformados en información; asimismo, considera el análisis de las situaciones de incertidumbre orientado a la toma de decisiones mediante el uso del enfoque clásico de probabilidad y los procedimientos asociados a su cálculo.

c. Contextos

Aluden a las condiciones que determinan la situación o la actividad propuesta. Pueden presentar o no conexiones con objetos o fenómenos de la realidad.

- *Intramatemático.* Es una situación que alude directamente a los objetos matemáticos. Se desarrolla exclusiva o principalmente en el plano de objetos, conceptos y procedimientos matemáticos abstractos. Las tareas presentadas aquí requieren para su solución procesos de matematización vertical (Treffers, 1987); es decir, se parte de un estado inicial de contexto puramente matemático y se le transforma, sin conectarlo con objetos o fenómenos de la realidad, hasta llegar al estado final deseado.
- *Extramatemático.* Es una situación que alude directamente a objetos y situaciones simuladas de la realidad, por lo que se le presenta al estudiante la tarea en un entorno que puede ser personal, familiar o comunal. En este caso, los conceptos son presentados como

útiles matemáticos, asociados generalmente a interpretaciones y convenciones culturales. Las tareas presentadas aquí requieren para su solución tanto procesos de matematización horizontal como vertical (Treffers, 1987); es decir, se parte de un estado inicial de contexto real que luego se transforma en un problema dentro del mundo matemático para llegar a los resultados matemáticos requeridos —es en esta fase en la que los resultados matemáticos deben ser interpretados a la luz de la situación planteada— y, finalmente, se transita ya desde el mundo matemático al mundo real para verificar la conveniencia y viabilidad de la respuesta hallada.

7.4 Diseño de la prueba

En esta sección, se presenta la distribución de ítems por cada aspecto del modelo de evaluación en Matemática.

a. Capacidades

En la siguiente tabla, se muestra cómo se distribuyen los ítems según la capacidad que evalúan.

Tabla 7.1. Matemática. Distribución de ítems de la prueba por capacidad a evaluar

Capacidad	Cantidad de ítems	
	n	%
Matematiza situaciones.	35	38,9
Comunica y representa ideas matemáticas.	25	27,8
Elabora y usa estrategias.	16	17,8
Razona y argumenta generando ideas matemáticas.	14	15,6
TOTAL	90	100,0

La prueba presta especial atención a la capacidad de matematización debido a las potencialidades que ofrece el lenguaje simbólico (particularmente algebraico), y la habilidad de generalización y abstracción que desarrollan los estudiantes del ciclo evaluado. Así, ellos pueden interpretar, elaborar o utilizar modelos sencillos de fenómenos

del mundo real que involucran relaciones matemáticas, por ejemplo de proporcionalidad, igualdad o paralelismo.

b. Contenidos

A continuación, se presenta la manera como se distribuyen los ítems de la prueba de Matemática de acuerdo con el contenido que evalúan.

Tabla 7.2. Matemática. Distribución de ítems de la prueba por contenido

Capacidad	Cantidad de ítems	
	n	%
Cantidad	23	25,6
Regularidad, equivalencia y cambio	27	30,0
Forma, movimiento y localización	22	24,4
Gestión de datos e incertidumbre	18	20,0
TOTAL	90	100,0

En el ciclo evaluado se priorizan los aprendizajes vinculados al tratamiento de las regularidades, la equivalencia y el cambio, y se supera la tradicional primacía que tenía el ámbito temático de cantidad. Esto es conveniente porque converge, por un lado, la importancia de la elaboración de conceptos y procedimientos asociados al cambio, el manejo de equivalencias y el tratamiento más exigente de las regularidades con un sentido funcional, y, por otro, las posibilidades de los estudiantes para darles una atención sistemática gracias a su mayor capacidad de abstracción, la apropiación creciente del lenguaje algebraico y el acceso a un universo numérico más potente (los números racionales).

c. Contextos

A continuación, se presenta la distribución de ítems atendiendo al contexto de las preguntas.

Tabla 7.3. Matemática. Distribución de ítems de la prueba por contexto

Contexto	Cantidad de ítems	
	n	%
Intramatemático	9	10,0
Extramatemático	81	90,0
TOTAL	90	100,0

Características de la construcción de las pruebas

8.1 Tipos de ítems

a. Ítems de opción múltiple y respuesta construida

En la teoría y práctica de la evaluación, particularmente aquella de lápiz y papel, se ha hecho una distinción entre dos de los formatos más recurrentes de ítems: los de opción múltiple y los de respuesta construida. Los ítems de opción múltiple, en general, presentan un enunciado y un número determinado de alternativas, de entre las cuales una es la correcta. Los ítems de respuesta construida requieren que el estudiante escriba la respuesta a la pregunta o tarea formulada en el enunciado (Osterlind, 1998).

Tradicionalmente, los cuestionarios usados para medir aprendizajes a gran escala han contenido únicamente preguntas de opción múltiple, con el fin de abaratar costos y eliminar la subjetividad de la calificación. Sin embargo, aunque las preguntas de opción múltiple pueden medir gran cantidad y diversidad de aprendizajes, hay también un grupo de habilidades que pueden enriquecer la descripción del proceso del aprendizaje y la validez, pero cuya medición enfrenta algunas limitaciones por el tipo de formato. Estudios como el de Ozuru, Briner, Kurby, y McNamara (2013) ofrecen evidencia sobre la calidad de la comprensión medida por las preguntas de respuesta construida frente a las de opción múltiple. Lissitz, Hou, y Slater (2012) señalan que “ambos tipos de ítems no necesariamente proveen la misma información ni elicitan la misma expresión de habilidades que pueden ser especialmente importantes cuando la prueba es usada para propósitos diagnósticos” (2012, p. 2). Estas observaciones sugieren que los ítems de respuesta construida, cuando menos, enriquecen la evaluación del aprendizaje.

Actualmente, los sistemas de evaluación alrededor del mundo han suscrito las limitaciones del uso exclusivo de ítems de opción múltiple y se han desarrollado procedimientos para incorporar y calificar respuestas construidas (Osterlind, 1998). Estos procedimientos son múltiples y rigurosos, y buscan que la calificación de este tipo de preguntas sea confiable (McClellan, 2010).

b. Ítems con créditos parciales

Por otro lado, es lógico considerar que las preguntas (ya sean de opción múltiple o de respuesta construida) no se resuelven necesariamente en un solo paso, sino que su resolución puede implicar un conjunto de etapas de procesamiento. Un estudiante que no ha resuelto la tarea completa puede haberse detenido en algún punto del proceso. En el pasado, las evaluaciones de sistema difícilmente eran sensibles a este aspecto de la evaluación. Hoy en día, se cuenta con modelos psicométricos que permiten incluir puntuaciones intermedias o “créditos parciales” que puedan informar de las habilidades que están en el camino hacia el logro de las tareas (Bond y Fox, 2007).

En la ECE, en las pruebas de Lectura y Matemática, se incluyen preguntas tanto de opción múltiple como de respuesta construida, así como ítems que contienen créditos parciales en su diseño, con el fin de hacer la prueba más sensible a habilidades complejas y reportar de mejor manera las habilidades “en camino” de las competencias evaluadas. Cabe señalar que se han construido créditos parciales tanto en ítems de respuesta construida como en los de opción múltiple. En este último caso, se ha asignado crédito parcial a la alternativa incorrecta que refleje un logro “en proceso” hacia la consecución de la tarea.

En la prueba de Escritura, los textos escritos por los estudiantes son calificados con rúbricas cuyos indicadores tienen dos o más créditos parciales.

c. Ítems con estímulo común

En las pruebas de Lectura y Matemática de la ECE, se considera, asimismo, la inclusión de conjuntos de ítems que se derivan de un mismo estímulo. Esta modalidad es frecuente en las evaluaciones de Lectura, donde el texto es el estímulo sobre el cual se diseñan ítems que miden diversos aspectos de la competencia lectora. Sin embargo, también se ha decidido su inclusión en la prueba de Matemática, puesto que, en la vida cotidiana, existen situaciones complejas que ofrecen la oportunidad de desplegar múltiples habilidades. De esta manera, se busca que la prueba se aproxime más a los contextos problemáticos reales.

8.2 Armado de las pruebas

a. Lectura y Matemática

En secundaria, los contenidos y las capacidades que deben lograr los estudiantes amplían y profundizan los aprendizajes logrados en la primaria. En Lectura, el lenguaje de los textos se complejiza y se especializa progresivamente, los temas son más diversos y las capacidades de inferencia y reflexión se consolidan. En Matemática, las nociones de número, variación y espacio, en un inicio ligadas al mundo concreto, requieren en secundaria mayor formalización, abstracción y el aprendizaje de conceptos nuevos y necesarios para seguir construyendo el pensamiento matemático.

Por esta razón, es necesario concebir una evaluación que cubra un amplio espectro de capacidades y conocimientos, de manera que sus resultados sean una fotografía más precisa de lo que realmente pueden hacer los estudiantes. Sin embargo, las evaluaciones de carácter censal suelen enfrentar la restricción de medir con una prueba única (y las consiguientes limitaciones del tiempo de evaluación y el número de ítems) las habilidades de los estudiantes.

Para lidiar con esta restricción, se estudió la posibilidad de utilizar diseños que contengan bloques de ítems comunes a todos los estudiantes y bloques no comunes que se distribuirían aleatoriamente entre la población. De esta manera, y gracias al uso de técnicas de equiparación desde el modelo Rasch, se podría contar con una mayor cantidad de ítems en la misma métrica, sin que esto implique incrementar la cantidad de ítems a los que cada estudiante se enfrenta.

Con la asistencia técnica del Berkeley Evaluation and Assessment Research (BEAR) Center de la Universidad de California (Berkeley), se condujo un conjunto de estudios psicométricos en el que se plantearon diversos escenarios que combinaban bloques comunes, bloques rotados y número de ítems por bloque. A partir de estos escenarios propuestos, se realizaron simulaciones estadísticas con la base de datos de la piloto para verificar que la confiabilidad de las personas en estos diseños se mantuviera estable en múltiples replicaciones, de manera que sea factible, dentro de márgenes razonables, reportar resultados a nivel individual y 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 y el valor estimado (entre 0,95 y 0,96).

Según estas configuraciones, en Lectura la escala de dificultad se construye con aproximadamente 70 ítems, de los cuales cada estudiante se enfrenta a 40 (20 por cada día de aplicación). En Matemática, el total de ítems asciende a alrededor de 90, de los cuales cada estudiante se enfrenta a 50 (25 por cada día de aplicación).

El bloque común, en ambos casos, tiene el fin de optimizar la equiparación de las medidas y es construido asegurando ítems que reflejen la dispersión de las medidas a lo largo de la escala de dificultad.

b. Escritura

En Escritura, se aplica un único cuadernillo que está conformado por cuatro consignas con distinto grado de dificultad. Cada consigna es evaluada por codificadores expertos mediante una rúbrica analítica, la cual contempla un conjunto de criterios cuyo contenido ha sido explicado en el capítulo 5.

Modelo de medición

En evaluaciones a gran escala, es imperativo caracterizar el aprendizaje de los estudiantes con precisión. La Teoría Clásica de los Test (TCT), un modelo basado en la sumatoria de las puntuaciones obtenidas en cada pregunta, suponía una relación monotónica lineal (Muñiz, 1996) entre los puntajes directos y la cantidad de rasgo latente; es decir, a cantidades iguales de aumento de puntaje corresponden similares aumentos en la cantidad de rasgo latente. Esta correspondencia es usada ampliamente en investigación; sin embargo, no es demostrable. Los puntajes directos arrojan como resultado un nivel ordinal de medición (Wright y Linacre, 1989) y no dan lugar a una escala de intervalo.

Por ello, en evaluaciones de sistema suelen usarse modelos de las Teorías de Respuesta al Ítem (TRI) o bien los modelos Rasch, los cuales no suponen a priori la existencia de un nivel de intervalo en las puntuaciones directas, sino que los transforman de manera que dicha propiedad pueda ser demostrable en la nueva escala. En la ECE, se utiliza este último para el análisis de los ítems de las pruebas.

9.1 El modelo Rasch

El modelo parte del supuesto de que los resultados de la interacción entre personas e ítems no pueden estar totalmente predeterminados, sino que estos siempre implican un elemento de impredecibilidad (Wright, 1989). Esto conlleva al requerimiento de que, en términos probabilísticos, mientras mayor habilidad, mayor probabilidad de responder correctamente a un ítem; mientras más difícil un ítem, menos probable para cualquier persona acertarlo. Para ello, se establece un modelo matemático de tipo probabilístico que vincula la habilidad o rasgo latente de una persona con la probabilidad de respuesta correcta a un ítem.

Así, se establece que la probabilidad de respuesta de una persona ante un ítem es una función de la diferencia entre la medida de habilidad de esa persona y la medida de la dificultad del ítem.

En los modelos Rasch, la habilidad de las personas y las dificultades de los ítems se ubican en la misma métrica. Ello permite combinar la habilidad

de la persona y la dificultad del ítem para predecir el desempeño de una persona en un ítem cualquiera e identificar respuestas inesperadas. La idea central del análisis Rasch es construir una escala conformada por los ítems ordenados según su dificultad. Ello implica que, a mayor habilidad, la persona tendrá mayor probabilidad de acertar a los ítems y, por lo tanto, mayor número de respuestas correctas.

En la línea que se señaló en el capítulo 7, el diseño de la ECE permite contar con ítems dicotómicos e ítems que tienen créditos (o puntajes) parciales. El análisis Rasch permite modelar ambos tipos de ítems. A continuación, se describe brevemente cómo lo hace.

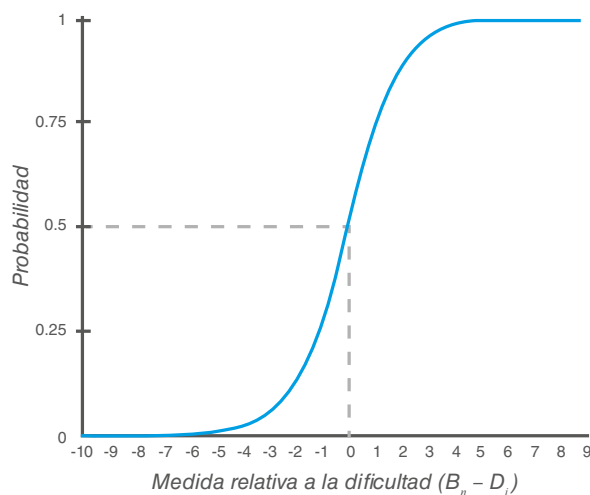
c. Modelo Rasch para ítems dicotómicos

Un ítem es dicotómico cuando solo tiene dos posibilidades de respuesta: error o acierto, y constituye quizás la forma más difundida de calificar ítems en las pruebas de lápiz y papel. La forma más simple del análisis Rasch ha sido diseñada para pruebas con ítems dicotómicos. El modelo expresa que la probabilidad de responder correctamente un ítem (obtener 1 antes que 0) es una función de la diferencia de la habilidad de la persona y la dificultad del ítem (Bond y Fox, 2007; Wright y Mok, 2004; Wright, 1977).

$$P_{-ni}(x=1)=f(B_{-n}-D_{-i})$$

La probabilidad de acertar el ítem puede graficarse mediante una curva característica del ítem que tiene la siguiente forma:

Gráfico 9.1. Curva característica para un ítem dicotómico



El gráfico informa que, a mayor diferencia entre la habilidad y la dificultad del ítem, mayor probabilidad de responderlo correctamente. Además, cuando la diferencia entre la habilidad de la persona y la dificultad del ítem es 0 (es decir, cuando ambas medidas son iguales), la probabilidad de responder correctamente es de 0,5.

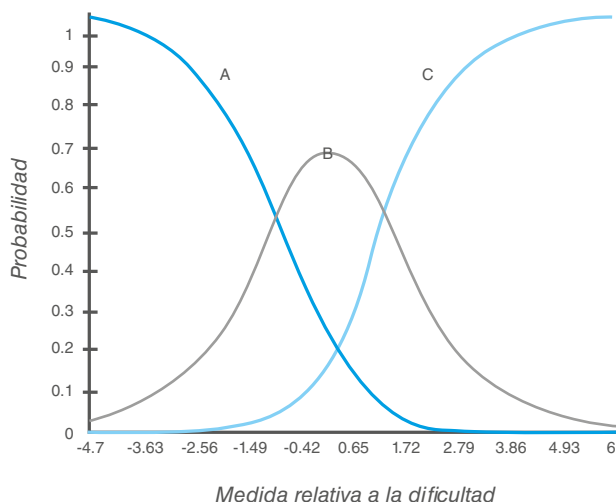
d. Modelo Rasch para ítems con créditos parciales

Como se ha visto, no solo el logro sino también el proceso de resolución de una tarea brindan información importante que aporta en la comprensión del constructo evaluado. Es lógico considerar que el estudiante, aún sin haber llegado al final de una tarea compleja, haya conseguido alcanzar algunos hitos en el proceso de resolverla. Es fundamental que la evaluación sea sensible a estos procesos; por ello, algunas preguntas de la prueba admiten respuestas que tienen puntajes o créditos parciales.

El análisis Rasch permite modelar este tipo de respuestas y establece en cada ítem umbrales que delimitan categorías en las que existe la probabilidad de conseguir una respuesta parcial antes que un error, o una respuesta adecuada antes que una parcial (Bond y Fox, 2007).

La probabilidad de responder en una de las categorías de un ítem con crédito parcial puede graficarse como sigue:

Gráfico 9.2. Curva característica para un ítem con un crédito parcial



En el gráfico, se presenta el caso de un ítem que tiene tres posibles respuestas: errada o inadecuada (con puntaje 0), parcialmente adecuada (con puntaje 1) y adecuada (con puntaje 2). La curva A representa la probabilidad de obtener 0 antes que 1; la curva B, la probabilidad de obtener 1 antes que 2; y la C, la probabilidad de obtener 2. Como puede observarse, en los niveles más bajos de la escala de habilidad, la categoría A es la que ocurre con mayor probabilidad; en los niveles intermedios de la escala, la categoría B; y en los niveles superiores, la C.

En la ECE, este modelo se utiliza para analizar los ítems de respuesta construida en los que se especifican más de dos códigos, así como aquellos de opción múltiple en los cuales algún distractor es considerado como una respuesta parcial.

9.2 Ajuste al modelo

En el proceso de construcción de un instrumento que mida adecuadamente el constructo deseado, es importante conocer en qué medida los datos recogidos por este reflejan lo que el modelo de medición prescribe. El modelo Rasch propone dos índices basados en estadísticos chi cuadrado de ajuste: el *outfit* y el *infit* (Linacre y Wright, 1994).

- *Outfit*. Está basado en el estadístico convencional chi cuadrado. Es particularmente sensible a patrones de respuesta inesperados por parte de personas frente a ítems muy fáciles o muy difíciles (Linacre y Wright, 1994).
- *Infit (inlier-pattern-sensitive fit statistic)*. Está basado en un estadístico chi cuadrado ponderado por la información estadística en cada observación. Es sensible a patrones de respuesta inesperados por parte de personas frente a ítems que están muy cercanos a la habilidad de la persona.

En la ECE, se admiten para el análisis aquellos ítems cuyos valores de *infit* y *outfit* se encuentren entre 0,7 y 1,3 (Linacre y Wright, 1994).

9.3 Funcionamiento diferencial de los ítems

Al diseñar una evaluación, se espera que el instrumento recoja la información de tal manera que las medidas reflejen la habilidad real del estudiante de la forma más objetiva posible. En ese sentido, los sistemas de evaluación deben procurar que los instrumentos eviten sesgos en la

construcción de los ítems que puedan facilitar o dificultar su resolución para un grupo en particular, por ejemplo, hombres frente a mujeres o poblaciones urbanas frente a rurales.

El modelo de evaluación supone que las personas con habilidades similares deberían tener similares probabilidades de responder al mismo ítem. Sin embargo, cuando, controlando la habilidad, existe una diferencia y esta es marcada y sistemática para un estrato, se dice que el ítem tiene funcionamiento diferencial (DIF, por sus siglas en inglés, *Differential Item Functioning*) (Zwick, Thayer, y Lewis, 1999). La presencia de DIF en un ítem puede advertir sobre posibles sesgos que hacen que un ítem sea más difícil para un grupo (por ejemplo, para los estudiantes de zona rural) que para otro (por ejemplo, los de zona urbana).

En la ECE, se considera primero si el DIF es estadísticamente significativo y luego se establecen dos categorías: a) un ítem tiene comportamiento diferencial de ligero a moderado si el valor absoluto del DIF es mayor o igual a 0,43 logits, y b) es moderado a grande si es mayor o igual a 0,64 logits (Linacre, 2015; Zwick et al., 1999).

Bibliografía

- Adam, J.-M. (2001). *Les textes: types et prototypes*. Armand Colin.
- AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association.
- Barton, D., & Hamilton, M. (2000). Literacy Practices. In D. Barton, M. Hamilton, & R. Ivanic (Eds.), *Situated Literacies. Reading and Writing in Context* (pp. 7–14). Londres: Routledge.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the Human Sciences* (2da ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Brennan, R. L. (2001). An Essay on the History and Future of Reliability from the Perspective of Replications. *Journal of Educational Measurement*, 38(4), 295–317.
- Britt, M. A., Goldman, S. R., & Rouet, J.-F. (2013). *Reading. From words to multiple texts*. New York NY: Routledge.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and Validity Assessment*. Beverly Hills, California: SAGE Publications.
- Cassany, D. (2003). Aproximaciones a la lectura crítica: teoría, ejemplos y reflexiones. *Tarbiya. Revista de Investigación E Innovación Educativa*, (32), 113–132.
- Cassany, D., & Castellà, J. M. (2010). Aproximación a la literacidad crítica. *Perspectiva, Florianópolis*, 28(2), 353–374.
- Cooke, R. L. (2013). *The History of Mathematics: A Brief Course*. Hoboken, N.J.: Wiley. Retrieved from <http://rbdigital.oneclickdigital.com>
- Cubukcu, F. (2007). An Investigation of Reading Strategies Employed by Trainee Teachers. *GEMA Online Journal of Language Studies*, 7(2), 95–110. Retrieved from http://pkukmweb.ukm.my/~ppbl/Gema/page95_110.pdf
- Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C., & Bivens-tatum, J. (2008). *Cognitive Models of Writing: Writing Proficiency as a Complex Integrated Skill*. New Jersey. Retrieved from <https://www.ets.org/Media/Research/pdf/RR-08-55.pdf>

- Flower, L., & Hayes, J. R. (2014). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365–387.
- Garner, R. (1987). Metacognition and reading comprehension. *Cognition and literacy*. Westport, CT, US: Ablex Publishing.
- Hanna, E. I. (2005). Inclusive Design for Maximum Accessibility: A Practical Approach to Universal Design. *Pearson Educational Measurement*, (August). Retrieved from http://www.pearsonassessments.com/NR/rdonlyres/BB1BC770-BCC1-4F06-9DA5-8D19A81E6C49/0/RR_05_04.pdf
- Hayes, J. R. (2000). A New Framework for Understanding Cognition and Affect in Writing. In R. J. R. Squire (Ed.), *Perspectives on writing: Research, theory, and practice* (pp. 6–44). Newark, DE, US: International Reading Association.
- Kintsch, W., y Mangalath, P. (2011). The construction of meaning. *Topics in Cognitive Science*, 3, 346-370.
- Kucer, S. B. (2005). *Dimensions of Literacy. A Conceptual Base for Teaching Reading and Writing in Schools* (2da ed.). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Linacre, J. M. (2014). Winsteps (Versión 3.81.0) [Manual y Software de computación]. Retrieved from <http://www.winsteps.com>
- Linacre, J. M. (2015). A user's guide to WINSTEPS, MINISTEP Rasch-Model computer programs. Retrieved from <http://www.winsteps.com/a/Winsteps-ManualPDF.zip>
- Linacre, J. M., & Wright, B. D. (1994). Chi – square fit statistics. In J. M. Linacre (Ed.), *Rasch Measurement Transactions Part II* (pp. 360–361). Chicago: MESA Press.
- Lissitz, R. W., Hou, X., & Slater, S. C. (2012). *The Contribution of Constructed Response Items to Large Scale Assessment. Journal of Applied Testing Technology*, 13(3), 1–52.
- McClellan, C. A. (2010). Constructed-Response Scoring — Doing It Right. *R&D Connections*, (13). Retrieved from https://www.ets.org/Media/Research/pdf/RD_Connections13.pdf

- Merzbach, U. C., & Boyer, C. B. (2011). *A History of Mathematics*. Hoboken, N.J.: Wiley.
- Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher*, 18(2), 5–11. <http://doi.org/10.3102/0013189X018002005>
- Messick, S. (1990). *Validity of Test Interpretation and Use*. Princeton, N. J.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <http://doi.org/10.1037/0003-066X.50.9.741>
- Meyer, A., Rose, D. H., & Gordon, D. (2013). *Universal Design for Learning: Theory and Practice*. Malden, MA: CAST.
- Ministerio de Educación - Oficina de Medición de la Calidad de los Aprendizajes. (2009). Marco de Trabajo de la Evaluación Censal de Estudiantes. Segundo grado de primaria y Cuarto grado de primaria EIB. Retrieved from http://umc.minedu.gob.pe/wp-content/uploads/2014/07/Marco_de_Trabajo_ECE.pdf
- Ministerio de Educación del Perú. (2015a). Rutas del Aprendizaje. Comunicación (Versión 2015). ¿Qué y cómo aprenden nuestros estudiantes? Lima: Autor. Retrieved from <http://recursos.perueduca.pe/rutas/secundaria.php#>
- Ministerio de Educación del Perú. (2015b). Rutas del Aprendizaje. Matemática (Versión 2015). ¿Qué y cómo aprenden nuestros estudiantes? Lima: Autor. Retrieved from <http://recursos.perueduca.pe/rutas/secundaria.php#>
- Muñiz, J. (1996). *Teoría Clásica de los Tests*. Madrid: Ediciones Pirámide.
- National Center On Universal Design for Learning. (n.d.). Universal Design for Learning Guidelines. Retrieved August 5, 2015, from <http://www.udlcenter.org/aboutudl/udlguidelines>
- National Council of Teachers of Mathematics. (2000a). *Principios y estándares para la educación matemática*. Sevilla: Sociedad Andaluza de Educación Matemática Thales.
- National Council of Teachers of Mathematics. (2000b). *Principios y estándares para la educación matemática*. Resumen Ejecutivo. Retrieved from

[http://www.nctm.org/uploadedFiles/Math_Standards/Executive Summary_Spanish_e-Final.pdf](http://www.nctm.org/uploadedFiles/Math_Standards/Executive_Summary_Spanish_e-Final.pdf)

- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. (J. Pelligrino, N. Chudowsky, & R. Glaser, Eds.). Washington DC: National Academy Press.
- Osterlind, S. J. (1998). *Constructing Test Items: Multiple-Choice, Constructed Response, Performance, Other Formats* (2nd ed.). New York NY: Kluwer Academic Publishers.
- Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 67(3), 215–227. <http://doi.org/10.1037/a0032918>
- Perfetti, C. A., & Roth, S. (1981). Some of the interactive processes in reading and their role in reading skill. *Interactive Processes in Reading*, 269–297.
- Quantitative Literacy Design Team. (2006). The Case for Quantitative Literacy. In B. Simeone & F. Pukelsheim (Eds.), *Mathematics and Democracy* (pp. 1–30). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Rumelhart, D. E., & McClelland, J. L. (1981). Interactive Processing Through Spreading Activation. In C. A. Perfetti & A. M. Lesgold (Eds.), *Interactive Processes in Reading* (pp. 37–60). New Jersey: Lawrence Erlbaum Associates.
- Story, M., Mueller, J. L., & Mace, R. L. (2011). The Universal Design File: Designing for People of All Ages and Abilities. *Design Research and Methods Journal*, 1(1), 165. Retrieved from <http://design-dev.ncsu.edu/openjournal/index.php/redlab/article/view/102>
- The National Council on Education and the Disciplines. (2006). *Mathematics and Democracy*. (B. Simeone & F. Pukelsheim, Eds.). Berlin, Heidelberg: Springer Berlin Heidelberg. <http://doi.org/10.1007/3-540-35605-3>
- Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). Universal design applied to large scale assessments. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>

- Thompson, S. J., & Thurlow, M. L. (2002). Universally designed assessments: Better tests for everyone! (Policy Directions No. 14). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Treffers, A. (1987). Three Dimensions. *A Model of Goal and Theory Description in Mathematics Instruction*—The Wiskobas Project. Dordrecht: Springer Netherlands. <http://doi.org/10.1007/978-94-009-3707-9>
- Treviño, E., Valdés, H., Castro, M., & Costilla, R. (2010). *Factores asociados al logro cognitivo de los estudiantes de América Latina y el Caribe*. Santiago de Chile. Retrieved from <http://disde.minedu.gob.pe/xmlui/handle/123456789/1068>
- Werlich, E. (1976). *A Text Grammar of English*. Heidelberg: Quelle und Meyer.
- Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14(2), 97–116. Retrieved from internal-pdf://wright_rasch1977-3682191616/Wright_Rasch1977.pdf
- Wright, B. D. (1989). Useful Measurement through One-Step Item Banking. In *Rasch Measurement Transactions Part 1* (p. 24). Chicago: MESA Press.
- Wright, B. D., & Linacre, J. M. (1989). Differences between scores and measures. In J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1* (pp. 63–65). Chicago: MESA Press.
- Wright, B. D., & Mok, M. M. C. (2004). An Overview of the Family of Rasch Measurement Models. *Introduction to Rasch Measurement*, 1–24.
- Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement*, 36(1), 1–28.

Ministerio de Educación

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

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



PERÚ

Ministerio
de Educación