

PUBLICACIONES

Reporte Técnico de la  
Evaluación Censal de  
Estudiantes (ECE 2013)  
Segundo y Cuarto (EIB) de primaria

COMPETENCIA  
DESEMPEÑOS  
LECTURA  
MATEMÁTICA  
ECE  
APRENDIZAJE  
EDUCACIÓN

REPORTE TÉCNICO DE LA EVALUACIÓN  
CENSAL DE ESTUDIANTES (ECE 2013)

SEGUNDO Y CUARTO (EIB) DE PRIMARIA



PERÚ

Ministerio  
de Educación

**LILIANA MIRANDA MOLINA**

JEFA DE LA UNIDAD DE MEDICIÓN DE LA CALIDAD EDUCATIVA

**REPORTE TÉCNICO DE LA EVALUACIÓN CENSAL DE ESTUDIANTES (ECE 2013)  
SEGUNDO Y CUARTO (EIB) DE PRIMARIA  
LIMA 2014**

**Responsables de la elaboración del reporte:**

**Andrés Burga León**

Responsable de la coordinación general del presente reporte

**Tania Pacheco Valenzuela**

Coordinadora del Equipo de Evaluación

**Nérida Urcia Harlet**

Coordinadora del Equipo de Planificación y Monitoreo

**Sandra Flores Ari**

Responsable del Equipo de Análisis

**Yuriko Sosa Paredes**

Especialista del Equipo de Análisis

**Yoni Arámbulo Mogollón**

Especialista de Área de Comunicación

**Fernando Llanos Masciotti**

Subcoordinador del Equipo de Lenguas Originarias

**Participaron en la revisión del presente reporte:**

Humberto Pérez León, Carmen Carpio, Miriam Arias y Olimpia Castro

**Correctora de estilo:**

Lucía de los Milagros Córdova Martínez

**Diagramador:**

Nilton Nakaza Tamayoshi

# Reporte técnico de la Evaluación Censal de Estudiantes (ECE 2013)

## Segundo y Cuarto (EIB) de primaria

<b>1</b>	<b>Introducción</b>	4
<b>2</b>	<b>Construcción de las pruebas aplicadas en la ECE</b>	6
2.1	Proceso de construcción de los instrumentos	6
2.2	Construcción de los ítems y las pruebas	7
2.2.1	Criterios para la selección de los textos de Comprensión Lectora	8
2.2.2	Validación de expertos 1: UMC	9
2.2.3	Validación de campo: Aplicación Piloto	9
2.2.4	Análisis de las propiedades psicométricas y de la distribución de los ítems en la escala	9
2.2.5	Validación de expertos EBR, DIGEIBIR e IPEBA	10
2.3	Aplicación de campo y constantes de equiparación	10
2.4	Aplicación definitiva	10
<b>3</b>	<b>Población y muestra</b>	12
3.1	Población objetivo	12
3.2	Marco muestral	12
3.3	Muestra de control	13
3.3.1	Estratificación	14
3.3.2	El efecto de diseño	15
3.3.3	El tamaño de la muestra	15
3.3.4	Cálculo de pesos	17
3.3.5	Cobertura y análisis de posibles sesgos	18
3.4	Estimación de resultados	19
<b>4</b>	<b>Operativo de campo</b>	20
4.1	Organización del operativo	20
4.2	Proceso de selección y capacitación de aplicadores	22
4.3	Proceso de recolección de datos	23
<b>5</b>	<b>Propiedades psicométricas de las pruebas aplicadas</b>	25
5.1	El modelo rasch para ítems dicotómicos	25
5.2	El modelo rasch frente a otros modelos psicométricos	29
5.2.1	El modelo rasch frente a la teoría clásica de los test	29
5.2.1.1	Independencia de las persona y del test	29
5.2.1.2	Medición de intervalo	29
5.2.2	El modelo rasch frente a la teoría de respuesta al ítem	30
5.2.2.1	Prescripción frente a descripción	30
5.2.2.2	El segundo parámetro o discriminación	30
5.2.2.3	El tercer parámetro o pseudoadivinación	30
5.2.2.4	El problema de la suficiencia	30
5.3	Programa informático utilizado para el análisis RASCH	31
5.4	Análisis de ítems	31
5.5	Evidencias de la confiabilidad de las medidas	38
5.6	Evidencias de validez	46
5.6.1	Evidencias de validez referidas al contenido de las pruebas	46
5.6.2	Evidencias de validez referidas a la estructura interna de las pruebas	48
5.7	Equiparación de puntuaciones	49
5.8	Niveles de logro y presentación de resultados	55
<b>6</b>	<b>Referencias</b>	57

# 1. Introducción

En el año 2006, el Ministerio de Educación del Perú tomó la decisión de llevar a cabo evaluaciones de carácter censal a los estudiantes de los primeros grados de primaria. En tal sentido, la Evaluación Censal de Estudiantes (ECE) evalúa a los estudiantes de segundo grado de primaria y, en caso de que en las instituciones educativas (IE) se aplique el Programa de Educación Intercultural Bilingüe (EIB), evalúa a los estudiantes de cuarto grado de primaria.

La evaluación se realiza con el propósito de monitorear el desarrollo de las habilidades fundamentales de los estudiantes para que continúen aprendiendo a lo largo del ciclo escolar. Se espera que en los primeros grados de primaria los estudiantes consoliden el aprendizaje de la lectoescritura, lo que permitirá que desarrollen habilidades de mayor complejidad, sobre las que se asentarán sus posteriores aprendizajes. Del mismo modo, se espera que, en los primeros grados, los estudiantes hayan adquirido el dominio básico de algunos conceptos matemáticos fundamentales, como lo son la estructura aditiva y la comprensión del sistema decimal de numeración. En este sentido, los objetivos de la ECE son los siguientes (Ministerio de Educación del Perú, 2009):

- Identificar el nivel de logro en Comprensión lectora en que se encuentran los estudiantes evaluados de segundo grado de primaria.
- Identificar el nivel de apropiación de la noción de número y sus operaciones en el campo aditivo en que se encuentran los estudiantes evaluados de segundo grado de primaria, así como su habilidad de utilizar esta comprensión en formas flexibles para hacer juicios matemáticos y desarrollar estrategias útiles para resolver problemas.
- Identificar el nivel de logro en Comprensión lectora en que se encuentran los estudiantes evaluados de cuarto grado de primaria de las IE que desarrollan, efectivamente, Educación Intercultural Bilingüe, según la definición operativa utilizada para efectos de la ECE.
- Determinar la evolución del rendimiento de los estudiantes a lo largo del tiempo en las capacidades y grados evaluados.
- Devolver resultados a todos los actores involucrados en la tarea educativa para que tomen decisiones que mejoren la calidad de los aprendizajes de los estudiantes.

Para lograr estos objetivos, la Unidad de Medición de la Calidad Educativa del Ministerio de Educación del Perú llevó a cabo, entre el 12 y el 15 de noviembre de 2013, la Evaluación Censal de Estudiantes, contando con el Instituto Nacional de Estadística e Informática (INEI) como operador logístico.

Dicho operativo está dirigido a las IIEE de nivel primario de la EBR con cinco o más estudiantes matriculados en el grado a evaluar (2.º o 4.º grado). En segundo grado de primaria, se aplicaron dos cuadernillos de Comprensión lectora y dos de Matemática. En cuarto grado de primaria EIB, se aplicaron dos cuadernillos de Comprensión lectora en castellano como segunda lengua.

El proceso de la ECE comprende las actividades de diseño, construcción y aplicación de instrumentos, recolección, procesamiento y análisis de información, y elaboración y difusión de resultados. En cuanto al procesamiento y análisis de información, en general, este implica los siguientes pasos:

1. Depurar las bases de datos, de tal manera que se garantice la calidad de la información que será utilizada para entregar resultados globales e individuales de la ECE.
2. Calibrar los ítems aplicando el modelo Rasch dicotómico. En esta etapa, se eliminan del análisis aquellos ítems que no han mostrado un adecuado ajuste al modelo Rasch.
3. Analizar el funcionamiento diferencial de los ítems a lo largo del tiempo (denominado ítem drift) aplicando

métodos gráficos e inferenciales, usando como punto de referencia las medidas estimadas para los ítems comunes entre ambas ECE, en año anterior.

4. Equiparar los ítems de la actual ECE con la ECE del año anterior. Para ello, se calculan las constantes de equiparación, omitiendo aquellos ítems que en el paso anterior han mostrado un marcado funcionamiento diferencial entre años.
5. Una vez equiparadas las medidas de dificultad de todos los ítems, estas son utilizadas para estimar las medidas de habilidad de todos los estudiantes evaluados.
6. Considerando el valor de las medidas de dificultad de los ítems que marcan los dos puntos de corte de la ECE, se asigna a cada estudiante a un nivel de desempeño, según su probabilidad de responder correctamente un ítem con una dificultad análoga a la del que marca el punto de corte.
7. Calcular los pesos por estudiante requeridos en la muestra de control, considerando el estrato al cual pertenece cada estudiante.
8. Calcular el porcentaje de estudiantes en cada nivel de desempeño, tanto para los resultados nacionales como por estrato, además de su respectivo error estándar, considerando el diseño muestral complejo utilizado.

El objetivo del presente documento es detallar los procesos de construcción de las pruebas aplicadas en la ECE 2013, la población y muestra utilizadas, la forma cómo se organizó el operativo y las estrategias de análisis psicométrico aplicadas. Se recomienda que la lectura del presente reporte técnico se haga en conjunto con el de fundamentación de la ECE (Minedu, 2009), documento en el que está basado.

En el segundo capítulo, se describe el proceso de construcción de los instrumentos aplicados, señalando los criterios seguidos en la elaboración de los ítems, recojo de evidencias de validez vinculadas al contenido de los ítems, y proceso piloto y definitivo de recojo de información.

El tercer capítulo está enfocado en la descripción de la población objetivo y el marco muestral utilizado. Además se señala qué es la muestra de control, cómo fue definido su tamaño y los pesos utilizados en la estimación de los diversos parámetros derivados de dicha muestra. Dicho capítulo culmina con el análisis de los posibles sesgos por cobertura.

El cuarto capítulo contiene los aspectos vinculados al operativo de campo y describe la forma cómo se organizó la red administrativa, el modo cómo fueron seleccionados y capacitados los aplicadores, y la manera cómo se recolectaron los datos en el contexto de la ECE 2013.

Finalmente, el quinto capítulo describe el modelo Rasch utilizado para los análisis psicométricos. Se pone énfasis en la forma cómo se analizaron los ítems y se obtuvieron evidencias de confiabilidad y validez. Además, se describe cómo se realizó el proceso de equiparación de medidas y la forma cómo se presentan los resultados de los estudiantes evaluados.

## 2. Construcción de las pruebas aplicadas en la ECE

### 2.1 PROCESO DE CONSTRUCCIÓN DE LOS INSTRUMENTOS

Los procesos de construcción, validación y aplicación definitiva de los instrumentos de la Evaluación Censal de Estudiantes (ECE) se desarrollan durante un periodo de tres años. En el primer año, se construyen las pruebas y se realiza la aplicación piloto. En el segundo, se aplican los instrumentos para obtener las constantes de equiparación y, finalmente, en el tercero, se aplican las pruebas definitivas en el operativo que se conoce como Evaluación Censal de Estudiantes. En el siguiente diagrama de flujo, se presenta de forma esquemática la secuencia de estos procesos y, luego, se explica cada uno de ellos.

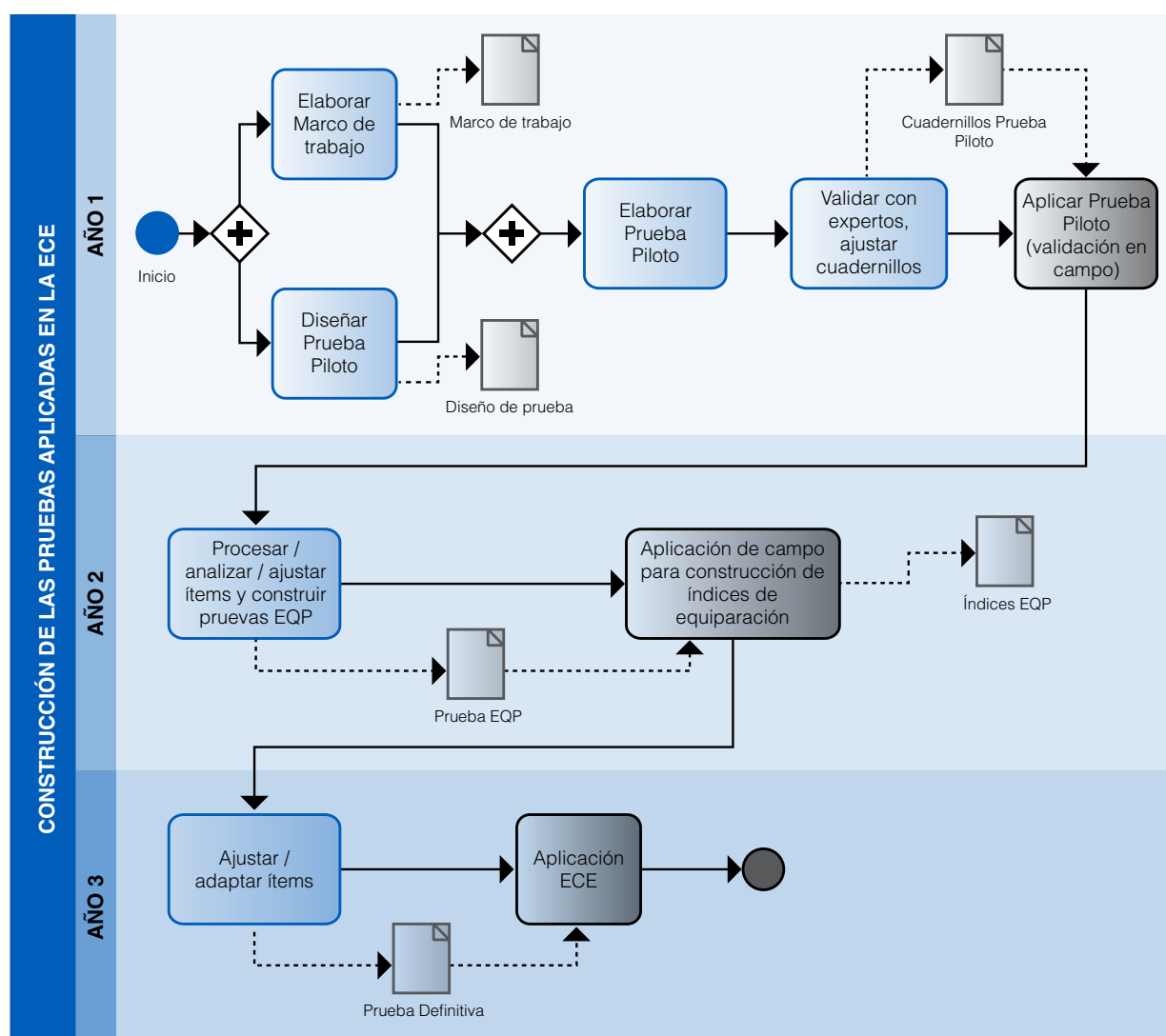


Figura 1. Diagrama de flujo con los procesos de la Evaluación Censal de Estudiantes

Según establece el *Marco de trabajo* (UMC, 2009), las pruebas de la ECE miden dos variables o constructos: comprensión lectora en Comunicación, y comprensión del número, de las operaciones y del sistema de numeración decimal en Matemática. Las capacidades implicadas en estos constructos, así como los indicadores y descriptores (considerados importantes o imprescindibles) que permiten evaluar dichas capacidades, se han plasmado desde

el año 2007 en tablas de especificaciones. En cada prueba, se emplean una cantidad de ítems que constituyen una muestra suficiente de ítems para medir los constructos de comprensión lectora y de comprensión del número, de las operaciones y del sistema de numeración decimal (46 y 42 ítems, respectivamente). La representatividad de los ítems empleados en las pruebas permite hacer inferencias acerca de la puntuación de los estudiantes en el universo de indicadores, a partir de la puntuación en la muestra de ítems que constituyen la prueba. Esto es posible gracias a que la muestra de ítems en las pruebas reproduce las características esenciales del universo de ítems relevantes destinados a medir los constructos, en la proporción y balance adecuados (McGartland, Berg-Weger, Tebb, Lee y Rauch, 2003; Muñiz, 1999; Suen, 1990).

En ese sentido, el proceso de validación de las pruebas de la ECE implica su mismo proceso de construcción, pues debe basarse en las tablas de especificaciones establecidas. De esta forma, se asegura que las pruebas contengan ítems que hagan referencia a cada una de las capacidades en una proporción adecuada. Por ello, los ítems empleados se definen en función del peso de las capacidades, indicadores y descriptores de las pruebas en la proporción y el balance adecuados (Muñiz, 1999). El hecho de que las pruebas de la ECE estén basadas en la tabla de especificaciones también ayuda a garantizar que los resultados obtenidos de las pruebas sean comparables en el tiempo. Esto quiere decir que, dado que año tras año se evalúan las mismas capacidades e indicadores mediante diferentes ítems, los instrumentos tienen mayores posibilidades de arrojar resultados comparables en el tiempo y así analizar los avances en el logro de los aprendizajes por parte de los estudiantes. Esto es de suma importancia, toda vez que retroalimenta al sistema educativo y permite tomar decisiones de política educativa basadas en datos válidos y confiables.

## 2.2 CONSTRUCCIÓN DE LOS ÍTEMS Y LAS PRUEBAS

Los instrumentos de la ECE son construidos por los especialistas de Matemática, Comunicación y Educación Intercultural Bilingüe del equipo de Evaluación de la UMC. Estos instrumentos son pruebas conformadas por ítems de *selección* o *elección múltiple*. En la estructura de un ítem, encontramos un *enunciado* o pregunta que se completa con tres *opciones de respuesta* o *alternativas* entre las que el estudiante debe identificar la única que es correcta. Esto implica que dos alternativas funcionan como *distractores*. Tanto en la construcción del enunciado como de las alternativas se sigue las pautas técnicas que la literatura ofrece. A continuación, se presenta las principales directrices en la construcción de los ítems ofrecidas por Moreno, Martínez y Muñiz (2004):

- A. Construcción del ítem en general
  1. Debe haber correspondencia entre el ítem y el indicador que evalúa.
  2. La complejidad de la tarea solicitada debe ser adecuada para el estudiante, según el grado.
  3. El ítem debe estar libre (en la medida de lo posible) de todo sesgo sociocultural y socioeconómico.
  4. El ítem no está relacionado con ningún otro ítem de la prueba, es decir, la respuesta correcta no puede inferirse a partir de otro ítem.
- B. Construcción del enunciado
  1. Lo central debe expresarse en el enunciado. Cada opción es un complemento de este.
  2. El enunciado debe concordar gramaticalmente con las alternativas.
  3. La sintaxis o estructura gramatical debe ser clara y precisa, es decir, debe estar claramente redactado. Se debe evitar enunciados demasiado escuetos, profusos, ambiguos o confusos.
  4. En lo posible, el enunciado no debe formularse de forma negativa. Si hubiera necesidad de usar una expresión negativa, debe resaltarse en mayúscula.
  5. El enunciado debe estar libre de elementos irrelevantes e innecesarios.
- C. Construcción de las alternativas
  1. La opción correcta debe ser solo una, acompañada por distractores plausibles.
  2. La opción correcta debe estar repartida entre las distintas ubicaciones.

3. Las opciones deben ser preferiblemente tres o cuatro. En la prueba se tiende a ofrecer tres, para evitar al niño un mayor esfuerzo cognitivo en la lectura, dada la edad que tienen.
4. Las opciones deben presentarse usualmente en forma vertical.
5. El conjunto de opciones de cada ítem debe aparecer estructurado.
6. Las opciones deben ser autónomas entre sí, sin solaparse ni referirse unas a otras.
7. Se debe evitar en todo momento las opciones «Todas las anteriores» y «Ninguna de las anteriores».
8. Ninguna opción debe destacar del resto ni en contenido ni en apariencia.
9. Las alternativas deben estar libres de palabras, frases o imágenes irrelevantes
10. Se debe emplear los errores posibles y más frecuentes en la resolución del ítem como distractores.

Tal como se pauta en la directriz C-10, los distractores deben construirse tomando en cuenta los errores más frecuentes que cometen los estudiantes al responder al ítem. En ese sentido, en la construcción de los distractores de los ítems de la ECE se utilizan las categorías de errores que se han encontrado en cada una de las dos áreas evaluadas. Estos errores frecuentes se presentan en la siguiente tabla.

**Tabla 1** Tipos de errores considerados para la elaboración de distractores en las pruebas ECE

Comprensión lectora	Matemática
Respuestas basadas en el saber previo vinculado al contenido del texto.	Respuestas que evidencian obstáculos típicos de los estudiantes ante los problemas propuestos (por ejemplo, sumar toda la información numérica del enunciado o identificar el mayor dato de los gráficos estadísticos).
Respuestas que muestran la utilización de información del texto de manera incorrecta.	Respuestas que dan cuenta de procedimientos parciales en la resolución de las tareas (atender solo a una de las etapas de una situación de varias etapas o indicar solo uno de los datos que intervienen en la resolución de la tarea propuesta).
Respuestas que dan cuenta de hipótesis no acomodativas, es decir, que el estudiante no logra acomodar sus hipótesis iniciales a la nueva información que va proporcionándole el texto, por lo que toma como respuesta la primera idea que se presenta en el párrafo (en ítems de tema central e idea principal).	Respuestas que evidencian dificultades en la jerarquización del Sistema de Numeración Decimal, mostrando una comprensión del número en términos de unidades solamente (en los ítems referidos a comparación u operaciones de adición y sustracción).
Respuestas que evidencian una jerarquización inadecuada de la información del texto.	Respuestas que muestran la tendencia a identificar palabras o frases que el estudiante considera claves para la elección de la estrategia a seguir (por ejemplo, «más», «más que», «en total», asociados a sumar; «menos», «quitar», asociados a restar).
Respuestas que muestran la confusión del estudiante sobre la información que es relevante para él (según su propósito) y lo que es importante para el lector.	
Respuestas que evidencian una lectura incompleta del texto.	

### 2.2.1 CRITERIOS PARA LA SELECCIÓN DE LOS TEXTOS DE COMPRESIÓN LECTORA

Un primer criterio es la diversidad de los contextos culturales que sirven de referente, tanto urbanos como rurales, así como elementos (objetos, ocupaciones, situaciones) de conocimiento universal. Se tiene un especial cuidado al ponderar estos aspectos para evitar posibles sesgos que perjudiquen a ciertos niños que, por su origen



o pertenencia a un estrato socioeconómico determinados, pudieran estar en desventaja respecto de otros. En el caso de cuarto grado EIB, en las pruebas de comprensión lectora en castellano como segunda lengua, dirigidas a niños de áreas rurales cuya lengua materna es originaria, se incluye en los textos temas de contextos rurales andinos y amazónicos. Por otro lado, en el caso de las pruebas de comprensión lectora en lenguas originarias, los textos refieren a temas de su propio contexto, pero sin dejar de lado textos de otras culturas.

Un segundo criterio es el de la diversidad de los tipos textuales utilizados. Según los enfoques actuales en lectura, la escuela debe contribuir a que, poco a poco, los niños se apropien de toda la variedad de textos necesaria para disfrutar, informarse de lo que sucede a su alrededor, estar preparados para aprender más y participar de las actividades de la comunidad o lugar donde viven. En concordancia con ello, los instrumentos ofrecen una amplia gama de tipos textuales (narrativos, expositivos, descriptivos, instructivos, argumentativos) en diversos géneros textuales de uso cotidiano (cartas, notas, afiches, artículos enciclopédicos, cuadros comparativos, recomendaciones, cuentos, anécdotas y noticias). De esta manera, se busca evaluar la lectura desde su uso funcional y social.

Un tercer criterio es el de la pertinencia del texto para el grado. En este aspecto, se ha controlado factores lingüísticos como la extensión, el vocabulario utilizado, las construcciones sintácticas y la densidad de información de cada texto. En el caso de cuarto grado EIB, en las pruebas de comprensión lectora en castellano como segunda lengua, se enfatizó aún más los aspectos gramaticales y de vocabulario considerando que el público evaluado son estudiantes que están en proceso de aprender el castellano como segunda lengua y que, probablemente, se encuentren en un nivel de dominio intermedio. Cabe señalar que tanto en segundo grado como en cuarto grado EIB se ha pretendido ofrecer textos de diversos niveles de complejidad, de manera que podamos tener una mayor variabilidad de resultados en la escala de dificultad.

### **2.2.2 VALIDACIÓN DE EXPERTOS 1: UMC**

Después de que los equipos de Matemática, Comunicación y de EIB han construido los ítems, estos son revisados por expertos de la UMC. Ellos evalúan aspectos como la calidad, actualidad 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 las alternativas, tanto en sus aspectos formales como en su eficacia para la medición del constructo. También se toman en cuenta posibles sesgos socioeconómicos, culturales y de género en la construcción. En el caso específico de Comunicación (segundo grado y cuarto grado EIB), se evalúa también la pertenencia de los textos a los tipos y géneros textuales que se señalan, la densidad de información, el vocabulario, la sintaxis, la familiaridad del tema, la extensión y el formato.

### **2.2.3 VALIDACIÓN DE CAMPO: APLICACIÓN PILOTO**

Después de la revisión de los expertos de la UMC, los ítems aprobados se utilizan para la construcción de las pruebas para la aplicación piloto, la cual consiste en una aplicación de las pruebas a una muestra representativa de estudiantes a nivel nacional. Las IE elegidas aleatoriamente corresponden a los estratos rural y urbano, estatal y no estatal, polidocente y multigrado / unidocente. Los estudiantes de esas IIEE presentan similares condiciones a las de la población objetivo (edad, avance curricular, desarrollo de capacidades, etc.). Además, las pruebas presentan similares estructura y características (orden de presentación de los ítems, según su dificultad; extensión, dificultad, diagramación, tiempo de aplicación de la prueba) a las de las pruebas definitivas. Todas estas condiciones de la aplicación piloto buscan asegurar que los datos encontrados permitan afirmar que los instrumentos tienen adecuadas evidencias de validez referidas a los constructos que evalúan.

### **2.2.4 ANÁLISIS DE LAS PROPIEDADES PSICOMÉTRICAS Y DE LA DISTRIBUCIÓN DE LOS ÍTEMS EN LA ESCALA**

Sobre la base de los resultados de la aplicación piloto, se analizan las propiedades y otros aspectos psicométricos de las pruebas y los ítems, así como una revisión pedagógica de la distribución de los ítems en la escala. Entre los aspectos psicométricos, se analiza el ajuste de los ítems y 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, sexo). Igualmente, se ensaya una hipótesis sobre las posibles causas por las que un ítem no ajustó al modelo psicométrico, tuvo un comportamiento diferenciado según los estratos o su dificultad no concordó con la complejidad estimada. Los valores estimados dan cuenta de la unidimensionalidad de las pruebas y son evidencia sólida de la validez de constructo de las mismas (en el capítulo 5 de este documento se especifican los detalles de las propiedades psicométricas de los ítems).

En lo que concierne a la revisión pedagógica, se analiza la correspondencia entre la complejidad pedagógica estimada de un ítem (antes de la piloto) y la dificultad dada por el modelo psicométrico (después de la piloto), la pertinencia de incluir un determinado ítem en la prueba definitiva, las posibles causas por las que los estudiantes no contestaron correctamente un ítem. En el caso específico de Comunicación, tanto en segundo grado como en cuarto grado EIB, se analizan también las posibles causas de error derivadas de algún factor relacionado con el texto. Este proceso de análisis concluye con la selección de los ítems de mejor desempeño y el ajuste de los ítems que presentaron algún tipo de problema.

### **2.2.5 VALIDACIÓN DE EXPERTOS EBR, DIGEIBIR E IPEBA**

Según lo establecido por la UMC, el proceso de validación incluye que un conjunto de expertos en las áreas evaluadas emita un juicio respecto de la construcción, la correspondencia con el indicador y la dificultad de la tarea pedida de cada ítem empleado para medir el constructo implicado en cada prueba (criterio de jueces). Por ello, los ítems seleccionados mediante la piloto y el análisis posterior se organizan en cuadernillos para que sean validados por expertos de la Dirección de Educación Primaria (DEP), de la Dirección General de Educación Intercultural Bilingüe y Rural (Digeibir) y del Instituto Peruano de Evaluación, Acreditación y Certificación de la Calidad de la Educación Básica (Ipeba). En esta segunda validación de expertos, se solicita a los jueces que evalúen los aspectos mencionados mediante una ficha técnica (véase anexos 4, 5, 6 y 7). Esta ficha tiene cuatro secciones: tres en que se presentan indicadores que evalúan las características generales del ítem, así como las características específicas del enunciado y de las alternativas, y una para que el juez escriba sugerencias puntuales sobre el ítem. Las fichas presentan ciertas particularidades, según el ítem esté incluido en las pruebas que evalúan los constructos de comprensión lectora o la comprensión del número, de las operaciones y del sistema de numeración decimal. En el caso específico de Comunicación, también se incluyen indicadores referidos al texto. Cada experto completa la ficha y formula las observaciones que, según 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, en Comunicación, al texto.

### **2.3 APLICACIÓN DE CAMPO Y CONSTANTES DE EQUIPARACIÓN**

Con los ítems aprobados por los expertos, se construyen las pruebas de equiparación. Estas se aplican en los dos días siguientes a la ECE del año anterior en un operativo denominado *Equiparación* (EQP), cuyos detalles técnicos se encuentran en la sección 5. Por ejemplo, en el año 2012, la ECE se realizó los días 4 y 5 de diciembre. En los dos días siguientes (6 y 7 de diciembre), se aplicaron las pruebas EQP. El objetivo de esta aplicación es calcular las constantes de equiparación. Este indicador es muy importante, pues permite convertir las medidas de las pruebas del año actual a las del año anterior. Este procedimiento es uno de los requisitos técnicos que garantizan la comparabilidad de los resultados a través del tiempo.

### **2.4 APLICACIÓN DEFINITIVA**

Finalmente, en el tercer año de todo este proceso se realiza la aplicación definitiva o más conocida como la ECE. Esto siempre ocurre en el último bimestre del año (noviembre-diciembre). Son dos días de aplicación. En el primer día, se aplica primero las pruebas de Comprensión lectora, luego, las de Matemática. El segundo día, el orden de aplicación es inverso. El cambio en el orden de aplicación, según el día, obedece a la necesidad de neutralizar los posibles efectos del cansancio de los niños al enfrentar las dos pruebas en un día. En el caso

de cuarto grado EIB, cuando se evalúa solo en castellano como segunda lengua, se aplica dos pruebas de Comprensión lectora en momentos diferentes durante un mismo día. En cambio, los años en que se evalúa en castellano como segunda lengua y en la lengua originaria de cinco comunidades lingüísticas: quechua Cusco-Collao, aimara, shipibo-conibo y awajún (a partir del 2014, se evaluará también en quechua chanca) se aplica en dos días. En el primer día, se aplica la prueba de Comprensión lectora en alguna de las lenguas originarias, luego en castellano como segunda lengua. El segundo día, el proceso es inverso.

En resumen, mediante los procesos de construcción y validación, se logra que los instrumentos de evaluación de la ECE 2013 cumplan con los requerimientos técnicos y consideraciones de contenido que permiten recoger datos válidos y confiables sobre los niveles de logro de nuestros estudiantes de segundo grado y cuarto grado EIB en comprensión lectora en Comunicación y comprensión del número, de las operaciones y del Sistema de Numeración Decimal en Matemática. Asimismo, los instrumentos guardan correspondencia con los diversos documentos curriculares del sistema educativo y sus resultados son comparables en el tiempo.

## 3. Población y muestra

### 3.1 POBLACIÓN OBJETIVO

La población objetivo de la Evaluación Censal de Estudiantes (ECE) son los estudiantes que asisten regularmente a clases en el segundo grado de primaria de una institución educativa (IE) de Educación Básica Regular (EBR) o estudiantes que asisten al cuarto grado de primaria en una escuela que desarrolla el Programa Educación Intercultural Bilingüe (EIB) según el director de la IE de EBR.

Se considera evaluar en segundo grado de primaria porque en este grado se concluye el tercer ciclo de la EBR. En este ciclo, se debe consolidar el aprendizaje de la lectoescritura, desarrollándose en forma simultánea las tareas lectoras que permiten el desarrollo de habilidades de mayor complejidad, sobre las que se asentarán los posteriores aprendizajes. Del mismo modo, se espera que en este ciclo los estudiantes hayan adquirido el dominio básico de algunos conceptos matemáticos fundamentales, como lo son la estructura aditiva y la comprensión del sistema decimal de numeración.

Atendiendo a esta distinción de que no todos los estudiantes manejan la lengua castellana, se considera evaluar a los estudiantes de cuarto grado de primaria de las instituciones educativas EIB, por ser este el grado en que finaliza el cuarto ciclo de la EBR y porque se espera que los estudiantes hayan concluido el aprendizaje inicial de lectoescritura (decodificación) tanto en lengua originaria<sup>1</sup> como en castellano como segunda lengua.

Exclusiones: se excluye a los estudiantes que asisten a una IE que atiende a menos de cinco estudiantes matriculados<sup>2</sup> en el grado a ser evaluado (segundo grado de primaria en lengua castellana o cuarto grado de primaria en lengua originaria). Esta exclusión es del 4% de estudiantes a nivel nacional, y con relación a la cantidad de IE es de alrededor del 27% de las excluidas a nivel nacional (Minedu, 2013). Las razones de esta exclusión se encuentran detalladas en el Marco de la ECE (Minedu, 2009).

En el año 2013, la ECE tuvo programado evaluar a 24 446 IE en segundo grado de primaria y 1 666 IE en cuarto grado de primaria; aproximadamente 560 000 estudiantes y 20 000 estudiantes en segundo y cuarto grado de primaria, respectivamente.

### 3.2 MARCO MUESTRAL

Está constituido por el listado de IE a ser evaluadas en la ECE y ha sido elaborado con información del Padrón de Instituciones Educativas de la Unidad de Estadística Educativa (UEE), la información de cantidad de estudiantes y secciones del Censo Escolar (CE) 2012 y del Sistema de Información de Apoyo a la Gestión de la Institución Educativa (Siagie) del año 2012.

Anualmente, la Unidad de Estadística Educativa (UEE) del Ministerio de Educación elabora y difunde un Padrón de IE con fines estadísticos, el cual es actualizado de manera constante. La información sobre IE tiene como fuente las operaciones estadísticas que aplica a nivel nacional y los procedimientos asociados a estas operaciones. Es importante precisar que la información de cantidad de estudiantes y secciones que proviene del CE es declarada por el director de la IE.

En la información del CE existen algunas IE que no reportan información, en las cuales la información corresponde a datos imputados, en el CE el 10% de los registros son imputados (un registro corresponde a una IE). Las características de las IE con información imputada son mayormente estatales (80%), rurales (59%) y pequeñas, es decir, menores a diez estudiantes (56%).

<sup>1</sup> Considerando que el aprendizaje de la lectoescritura en niños y niñas bilingües implica el conocimiento y manejo de una cultura escrita para quienes provienen de una cultura oral.

<sup>2</sup> La cantidad de estudiantes matriculados es el dato mayor entre lo registrado por el Sistema de Información de Apoyo a la Gestión de la Institución Educativa (Siagie), Censo Escolar y aplicación de la ECE del año anterior.

Para fines de elaborar el listado de IE de la ECE 2013, se siguieron los siguientes procedimientos:

1. Solicitar el Padrón de Instituciones Educativas a la UEE, la información de la cantidad de estudiantes y secciones del Censo Escolar 2012, y la cantidad de estudiantes y secciones del Siagie.
2. Revisar consistencias en la información solicitada.
3. Identificar las IE del Programa Educación Intercultural Bilingüe (EIB). Para ello, se utiliza la definición operativa construida para efectos de la ECE (Minedu, 2009), la cual sostiene que una IE es EIB si atiende a niños que tienen una lengua originaria y en ella se enseña a leer y escribir en dicha lengua y en castellano como segunda lengua, según la declaración del director de la IE. En la ECE 2013, se evaluó Comprensión lectora en castellano como segunda lengua. Se debe recordar que la evaluación en lengua originaria es interanual (Minedu, 2009).
4. Asignar el grado a evaluar. Las IE EIB son evaluadas en cuarto grado, y, las demás, en segundo grado.
5. Calcular la cantidad de estudiantes y secciones programadas. Esta información corresponde al valor mayor entre la información del Censo Escolar del año 2012, Siagie 2012 o estudiantes evaluados en la ECE 2012 dependiendo del grado a evaluar.
6. Seleccionar las IE que tienen cinco o más estudiantes asignados en el punto 5.

La distribución de IE y estudiantes en la ECE 2013 mediante el procedimiento antes descrito se presenta en la tabla 2.

**Tabla 2** Cantidad de IE, secciones y estudiantes programados en la ECE 2013

Tipo de evaluación	IE	Secciones	Estudiantes
Segundo grado	24 446	33 294	564 862
ECE 2013	18 821	23 850	371 052
MC 2013	5 625	9 444	193 810
Cuarto grado	1 666	1 756	20 590

### 3.3 MUESTRA DE CONTROL

Para garantizar resultados precisos en la ECE segundo grado<sup>3</sup>, se decidió reportar resultados a partir de una Muestra de Control (MC), con los siguientes niveles de inferencias: i) nivel nacional, ii) por género de los estudiantes, iii) por gestión de las IE (estatales / no estatales), iv) por ubicación geográfica de las IE (urbanas / rurales), v) a nivel de Dirección Regional de Educación (DRE). Las razones de usar la MC se detallan en el *Marco de trabajo de la ECE* (Minedu, 2009).

La MC es una muestra probabilística de instituciones educativas. Cada IE es considerada un conglomerado, pues agrupa a un conjunto de estudiantes, quienes son la unidad del análisis.

El muestreo de conglomerados surge como una necesidad práctica, económica y de eficiencia administrativa. Su ventaja más importante es que no es necesario contar con un marco muestral de cada elemento a analizar, solo es necesario tener el marco muestral a nivel de conglomerados (IE) para realizar la selección de los mismos (Lehtonen y Pahkinen, 2004). Luego, en cada conglomerado seleccionado, se puede obtener el listado de todos los estudiantes y seleccionar una muestra de estos. En la MC, todos los estudiantes forman parte de la muestra.

<sup>3</sup> En el caso de cuarto grado EIB, no era necesario hacer esto dado que el tamaño de la población a evaluar es pequeño.

### 3.3.1 ESTRATIFICACIÓN

En la MC, se realiza una estratificación principalmente por las siguientes razones:

- mejorar la precisión de los estimadores;
- aplicar alguna particularidad de diseño de muestra en algunos estratos;
- asegurar una adecuada representación de grupos específicos en la población;
- conveniencia administrativa.

La MC es estratificada por la Dirección Regional de Educación (DRE), bajo la consideración que las IE son administradas por estas instancias, y al interior de estas se forman estratos por tamaño de conglomerado (IE), que está determinado por la cantidad de estudiantes en la institución educativa. La distribución de las IE y estudiantes por estrato, se presenta en la tabla 3.

**Tabla 3** Porcentaje de IIEE y estudiantes, según estrato tamaño de la IE

DRE	[5-9]		[10-17]		[18-34]		[35 a más]	
	IIEE	Estudiantes	IIEE	Estudiantes	IIEE	Estudiantes	IIEE	Estudiantes
Amazonas	49,0	20,6	28,1	21,7	15,3	23,0	7,7	34,6
Áncash	34,2	11,1	32,7	19,7	19,5	20,8	13,6	48,3
Apurímac	37,1	12,5	32,9	21,3	14,9	17,8	15,1	48,4
Arequipa	24,4	7,0	25,8	14,4	26,8	25,7	22,9	52,9
Ayacucho	41,9	13,8	26,3	16,0	17,5	20,0	14,3	50,1
Cajamarca	50,7	23,3	28,8	24,1	13,7	21,2	6,7	31,4
Callao	15,5	3,6	22,2	8,0	28,4	18,6	34,0	69,9
Cusco	38,0	10,9	25,7	13,3	18,8	19,8	17,5	56,1
Huancavelica	47,2	20,7	30,3	25,0	15,2	23,2	7,3	31,1
Huánuco	42,2	16,0	29,2	20,2	17,9	23,8	10,7	40,0
Ica	21,3	4,7	21,5	8,8	25,9	17,7	31,4	68,8
Junín	41,7	13,2	29,1	18,2	16,7	19,1	12,5	49,5
La Libertad	31,0	8,5	26,2	13,3	23,7	22,2	19,1	55,9
Lambayeque	24,9	5,2	26,0	10,7	21,8	15,0	27,3	69,0
Lima Metropolitana	13,9	3,2	23,3	8,8	28,4	18,6	34,4	69,3
Lima Provincia	32,0	8,0	23,5	11,2	22,0	19,6	22,6	61,2
Loreto	43,0	11,9	26,3	13,6	14,1	13,8	16,6	60,7
Madre de Dios	35,9	6,5	17,2	7,0	22,7	17,2	24,2	69,3
Moquegua	36,8	6,7	13,7	5,8	20,0	15,2	29,5	72,4
Pasco	51,1	18,7	23,5	14,5	15,2	19,8	10,2	47,0
Piura	29,6	7,7	29,6	14,0	19,7	17,6	21,1	60,7
Puno	44,2	13,6	25,0	15,1	15,6	16,9	15,3	54,4
San Martín	35,9	10,7	26,9	13,9	20,1	21,2	17,2	54,2
Tacna	18,4	2,9	21,1	8,3	24,5	16,1	36,1	72,8
Tumbes	21,3	3,4	18,4	6,4	25,7	15,9	34,6	74,3
Ucayali	42,1	6,8	18,2	5,2	11,6	7,4	28,0	80,6
<b>TOTAL</b>	<b>33,7</b>	<b>8,9</b>	<b>26,5</b>	<b>13,3</b>	<b>20,3</b>	<b>19,0</b>	<b>19,6</b>	<b>58,8</b>

Como se observa en la tabla 3, el estrato más pequeño concentra la mayor cantidad de IE (34%); sin embargo, a nivel de estudiantes la contribución del estrato pequeño es mínima (de 9%).

En la MC se han definido dos tipos de estratos: explícitos e implícitos. Los estratos explícitos: DRE y tamaño, han sido definidos para tratar de manera independiente a las unidades en cada uno de estos estratos.

También se ha definido estratos implícitos con el fin de lograr una distribución proporcional de IE. Los estratos implícitos definidos para la MC son gestión de la IE, área de ubicación de la IE, característica de la IE, provincia y distrito de ubicación de la IE.

### 3.3.2 EL EFECTO DE DISEÑO

Para muestras complejas, se debe tener en cuenta el efecto de diseño. Kish (1965) describe el efecto de diseño como el ratio de la varianza del estimador obtenido de la muestra más compleja y la varianza estimada obtenida de la muestra aleatoria simple del mismo número de unidades. Cochran (1977) señala que el efecto de diseño tiene dos usos principales que son los siguientes:

- estimar el tamaño de muestra,
- evaluar la eficiencia de planes más complejos.

En el caso de la Muestra de Control se ha calculado el efecto de diseño para cada DRE con el fin de obtener los tamaños de muestra en cada una de ellas. La fórmula para el cálculo del efecto de diseño se presenta a continuación:

$$deff = 1 + \rho(m - 1) \quad (3.1)$$

En ella, *deff* es el efecto de diseño,  $\rho$  es la correlación intraclase, y *m* es el tamaño promedio del conglomerado.

El coeficiente de correlación intraclase es un indicador de la proporción de variabilidad del logro de aprendizaje entre conglomerados y se calcula con la siguiente fórmula:

$$\hat{\rho} = \frac{F-1}{F+(m-1)} \quad (3.2)$$

En la que *F* es el valor calculado en el análisis de varianza (Lord, 1999).

Cuanto mayor sea la correlación intraclase, los conglomerados son más similares al interior, pero diferentes entre conglomerados; por ello, es necesario tener más unidades en la muestra. Un ejemplo de ello se puede observar en el caso de la DRE Cusco: las instituciones educativas en esta región son muy diferentes entre ellas, pero homogéneas al interior. En el caso contrario, a menor correlación intraclase dentro de los conglomerados, las unidades son desiguales y entre conglomerados se genera cierta homogeneidad, por lo que se necesita menos conglomerados en la muestra. Este caso lo encontramos, por ejemplo, en Moquegua, donde las IE presentan mayor homogeneidad entre ellas.

### 3.3.3 EL TAMAÑO DE LA MUESTRA

El tamaño de muestra fue calculado considerando un intervalo de confianza al 95% para el promedio de la medida de habilidad en Comprensión lectora o el promedio de la medida de habilidad en Matemática de los estudiantes en segundo grado de primaria. La fórmula aplicada fue la siguiente:

$$n = deff \left( \frac{N\sigma^2 z^2}{(N-1)e^2 + \sigma^2 z^2} \right) \quad (3.3)$$

Donde los valores son los siguientes:

*n*= cantidad de estudiantes a muestrear

*N*= cantidad de estudiantes en la población

*z*= valor en el eje x de la distribución normal estándar cuando la probabilidad es 0.95

$\sigma$ = desviación estándar de la medida de logro en la población

*e*= error asumido para la muestra (amplitud del intervalo a estimar)

*deff*= efecto de diseño al utilizar un diseño por conglomerados

Para el cálculo de tamaño de muestra se ha considerado un error de estimación de 0,1 que genera un intervalo de confianza de amplitud 0,2 de la estimación del promedio de la medida de habilidad de los estudiantes en cada área (Comprensión lectora y Matemática). Luego se elige la mayor muestra de estudiantes necesarios y se calcula la cantidad requerida. Finalmente, considerando la proporción de estudiantes de cada estrato, las IIEE son distribuidas para los estratos de tamaño en cada una de las DRE. Las cantidades de muestra requeridas se presentan en la tabla 4.

**Tabla 4** Porcentaje de IIEE y estudiantes, según estrato tamaño de la IE

DRE	rho	deff	n	Cantidad de IIEE				
				[35 a más]	[18-34]	[10-17]	[5-9]	Total
Amazonas	0,3953	4,9	3 802	44	63	60	51	218
Áncash	0,4920	7,1	5 595	90	79	63	38	270
Apurímac	0,4679	7,1	4 445	72	58	58	32	220
Arequipa	0,2887	6,1	4 837	84	74	35	15	207
Ayacucho	0,4789	6,3	4 271	69	61	41	34	206
Cajamarca	0,4935	5,5	6 039	66	96	108	81	350
Callao	0,1795	4,5	3 920	87	34	18	6	145
Cusco	0,5487	8,4	6 120	108	78	54	36	276
Huancavelica	0,5404	6,2	4 366	53	65	78	53	249
Huánuco	0,5283	7,0	4 330	58	74	58	37	226
Ica	0,2127	5,0	4 447	94	46	21	9	170
Junín	0,4568	6,6	5 419	87	71	65	42	264
La Libertad	0,4324	7,3	5 483	98	74	46	23	241
Lambayeque	0,3932	7,3	5 102	109	48	27	14	198
Lima Metropolitana	0,1833	4,8	4 633	100	46	20	6	173
Lima Provincia	0,2861	5,3	4 210	81	47	29	17	175
Loreto	0,4845	6,8	4 752	93	40	44	34	211
Madre de Dios	0,3018	6,1	3 249	70	26	17	12	125
Moquegua	0,1348	2,7	2 425	54	16	17	8	95
Pasco	0,4171	5,5	3 329	53	41	34	40	168
Piura	0,4677	8,2	6 590	128	73	51	28	280
Puno	0,5012	7,6	4 547	80	50	45	36	212
San Martín	0,4615	6,8	5 024	86	64	49	31	231
Tacna	0,1852	4,5	3 017	68	25	11	5	109
Tumbes	0,2409	6,1	3 484	78	31	10	7	126
Ucayali	0,4357	7,0	4 663	110	24	21	24	179
<b>TOTAL</b>			<b>118 098</b>	<b>2 121</b>	<b>1 404</b>	<b>1 080</b>	<b>721</b>	<b>5 325</b>



### 3.3.4 CÁLCULO DE PESOS

Los pesos ayudan a corregir la distribución de la muestra en los estratos y también expanden la información muestral a la población. En esta etapa, se ha considerado una posestratificación por gestión, área geográfica y característica.

El peso es el inverso a la probabilidad de selección de un conglomerado (IE) al interior de cada estrato, es decir:

$$p_i = \frac{N_i}{n_i} \quad (3.4)$$

En esta fórmula, los valores son los siguientes:

$p_i$  = peso en el i-ésimo estrato

$N_i$  = instituciones educativas en el i-ésimo estrato

$n_i$  = instituciones educativas seleccionadas en el i-ésimo estrato

**Ajuste por instituciones educativas no evaluadas en el estrato.** Si una o más instituciones educativas no han sido evaluadas en un estrato, se realiza la siguiente corrección:

$$a1_i = \frac{n_i}{e_i} \quad (3.5)$$

Donde los valores son los siguientes:

$a1_i$  = ajuste por instituciones educativas no evaluadas en el i-ésimo estrato

$n_i$  = instituciones educativas seleccionadas en el i-ésimo estrato

$e_i$  = instituciones educativas evaluadas en el i-ésimo estrato

**Ajuste por estudiantes no evaluados en una institución educativa.** Esta corrección se realiza para cada área evaluada, es decir, para Comprensión lectora y Matemática de manera separada.

$$a2_{ij} = \frac{t_{ij}}{s_{ij}} \quad (3.6)$$

Donde los valores son los siguientes:

$a2_{ij}$  = ajuste por estudiante no evaluado en el i-ésimo estrato y la j-ésima IE

$t_{ij}$  = estudiantes que asisten al i-ésimo estrato y la j-ésima IE

$s_{ij}$  = estudiantes evaluados en el i-ésimo estrato y la j-ésima IE

**Ajuste por estudiantes no evaluados en el estrato.** Esta corrección se realiza para cada área evaluada, es decir, para Comprensión lectora y Matemática de manera separada.

$$a3_i = \frac{t_i}{s_i} \quad (3.7)$$

Donde los valores son los siguientes:

$a3_i$  = ajuste por estudiantes no evaluados del i-ésimo estrato

$t_i$  = estudiantes que asisten a la IE en el i-ésimo estrato

$s_i$  = estudiantes evaluados del i-ésimo estrato

Se obtiene un peso para Comprensión lectora y otro para Matemática. El peso final está dado por

$$pf_{ij} = p_i * a1_i * a2_{ij} * a3_i \quad (3.8)$$

Donde:

$pf_{ij}$  = peso final en el i-ésimo estrato y la j-ésima IE

$p_i$  = peso en el i-ésimo estrato

$a1_i$  = ajuste por instituciones educativas no evaluadas en el i-ésimo estrato

$a2_{ij}$  = ajuste por estudiante no evaluado en el i-ésimo estrato y la j-ésima IE

$a3_i$  = ajuste por estudiantes no evaluados del i-ésimo estrato

### 3.3.5 COBERTURA Y ANÁLISIS DE POSIBLES SEGOS

Es importante evaluar el rendimiento de la muestra para cada uno de los estratos definidos. Por ello, se ha calculado la cobertura para los estratos en los que se reportan los resultados al interior de cada DRE. La cobertura a nivel de estudiantes se muestra en la tabla 5. El porcentaje de estudiantes evaluados respecto de los programados a nivel nacional es del 90,7%. Dicha cobertura se calculó considerando la cantidad de estudiantes que llegan a la evaluación final del año escolar en cada institución educativa, según la información de Siagie para el año 2013.

**Tabla 5** Porcentaje de estudiantes evaluados por estratos en cada DRE

DRE	Gestión		Área		Característica	
	Estatal	No estatal	Urbana	Rural	Polidocente completo	Unidocente / Multigrado
Amazonas	92,8	90,7	94,6	87,7	94,6	85,8
Áncash	91,8	91,7	91,8	91,7	91,5	94,7
Apurímac	96,3	72,8	94,8	94,3	94,9	93,6
Arequipa	94,5	93,8	94,4	88,5	94,4	89,6
Ayacucho	86,4	90,4	87,3	84,1	86,4	91,0
Cajamarca	89,8	96,4	90,1	90,4	89,7	92,0
Callao	91,1	86,4	89,7	--	90,1	71,4
Cusco	94,8	97,2	95,9	91,8	95,5	92,3
Huancavelica	92,0	98,3	91,0	93,7	92,5	90,6
Huánuco	94,1	94,3	94,4	93,5	94,2	93,6
Ica	92,8	95,8	93,5	92,1	93,6	90,2
Junín	93,3	94,0	94,3	86,4	94,3	86,3
La Libertad	91,9	93,1	92,3	91,7	92,4	90,3
Lambayeque	92,5	93,3	93,1	89,0	92,9	85,6
Lima Metropolitana	87,9	89,6	88,6	--	88,8	83,3
Lima Provincia	87,8	89,7	88,0	92,5	88,3	88,8
Loreto	87,4	89,5	89,6	77,9	89,0	79,3
Madre de Dios	93,1	86,6	93,8	87,2	94,9	79,4
Moquegua	93,5	91,8	93,0	96,8	93,2	91,1
Pasco	93,2	98,0	95,2	86,1	94,9	83,3
Piura	89,8	88,9	88,8	94,9	89,3	91,8
Puno	96,8	97,2	97,0	95,7	97,0	93,5
San Martín	92,2	95,6	92,2	92,5	92,7	90,1
Tacna	95,4	94,2	94,9	97,3	95,4	81,3
Tumbes	85,2	86,6	85,3	87,9	85,8	72,8
Ucayali	82,9	93,8	83,4	81,9	83,4	82,1
<b>TOTAL</b>	<b>91,1</b>	<b>91,6</b>	<b>91,3</b>	<b>90,2</b>	<b>91,4</b>	<b>88,5</b>

Para el caso de las IE evaluadas en cuarto grado EIB, el porcentaje de estudiantes evaluados se presenta en la Tabla 6.

<b>Tabla 6</b> Porcentaje de estudiantes evaluados por lengua programada en EIB en cada DRE	
Lengua a evaluar en EIB	%
Aimara	96,3
Awajún	86,2
Quechua	95,5
Shipibo	79,0
Castellano como segunda lengua	88,0
<b>Total</b>	<b>88,7</b>

### 3.4 ESTIMACIÓN DE RESULTADOS

La estimación de resultados de la ECE en segundo grado es realizada a partir de la Muestra Control y se proporcionan dos tipos de resultados:

1. estimación del intervalo de confianza del promedio de la medida de habilidad obtenido por los estudiantes en cada uno de los niveles de inferencia definidos;
2. estimación del intervalo de confianza del porcentaje de estudiantes en cada nivel de logro alcanzado.

Los niveles de inferencia, como se menciona al inicio de esta sección, para los cuales se pueden reportar resultados son los siguientes:

- a nivel nacional,
- por género de los estudiantes: femenino, masculino,
- por gestión de la IE: estatal, no estatal,
- por área geográfica de la IE: urbano, rural,
- por característica de la IE: polidocente completo, unidocente / multigrado,
- por Dirección Regional de Educación (DRE).

Para el caso de la DRE de Madre de Dios, Moquegua, Tacna y Tumbes la estimación de resultados se realiza con todos los conglomerados al interior de estos estratos. La razón de esta estimación es debido a que se tienen pocos conglomerados y para no complicar la aplicación en estos estratos.

Los resultados son calculados utilizando el programa informático SPSS, específicamente, el módulo de muestras complejas. En este programa informático, la estimación de la varianza para una proporción (el ratio  $R = Y/X$ ) es aproximada usando la fórmula de linearización de Taylor (Woodruff, 1971).

Woodruff (1971) advierte que el método de cálculo de la varianza es completamente general y asume que la muestra es suficientemente grande para justificar el uso de la aproximación de Taylor, supone también que no se tiene restricciones en la forma del estimador, el número de variables aleatorias supuestas en el estimador, el tipo y complejidad o número de muestras diseñadas asumidas en el estimador. La muestra control en cada DRE, que es el estrato más pequeño para el cual se reportan resultados, es de tres mil a más estudiantes.

La estimación de resultados de cuarto grado para las instituciones educativas que son EIB se realiza de manera directa con la información de los estudiantes de las IE evaluadas en EIB.

## 4. Operativo de campo

### 4.1 ORGANIZACIÓN DEL OPERATIVO

Desde el año 2007, el Instituto Nacional de Estadística e Informática (INEI) es el operador logístico responsable de llevar a cabo la aplicación de los instrumentos de la ECE. Durante la ejecución de esta evaluación, el operador logístico y el Minedu implementaron controles de calidad de todos los procedimientos establecidos, los cuales quedaron plasmados en unos Términos de Referencia (TdR). El operador logístico se encarga de garantizar las siguientes tareas:

- la correcta impresión, modulado y embalaje de los instrumentos de evaluación, así como de los materiales de aplicación y capacitación;
- la distribución y entrega oportuna y segura de los materiales de capacitación y aplicación en las sedes provinciales y el retorno de los instrumentos aplicados a la empresa responsable de la captura de datos;
- la seguridad y confidencialidad requeridas 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 coordinador general y equipo directivo del INEI a cargo del operativo trabajan en conjunto con el Minedu para hacer el seguimiento de las consideraciones logísticas y pedagógicas señaladas en los Términos de Referencia (TdR).

El INEI lanzó dos concursos públicos, uno para la impresión y modulado de los instrumentos de la ECE; y otro para el transporte de los instrumentos. Como resultado de estos concursos, el operador logístico contrató a dos imprentas y dos empresas de transporte. El Minedu entregó a las imprentas contratadas por el operador logístico los archivos digitales en formato PDF de todos los instrumentos y materiales a imprimir:

- cuadernillos de pruebas ECE segundo grado de primaria y cuarto grado EIB, MC segundo grado de primaria, EQP segundo grado de primaria y cuarto grado EIB;
- papelógrafos con las indicaciones para la aplicación de las pruebas;
- manuales de aplicación para la RA.

En la ECE 2013, el operador logístico supervisó el avance y realizó el control de calidad de la impresión y modulado de los instrumentos. El Minedu monitoreó estos procesos. Se embolsó 36 487 cajas con instrumentos (una caja por sección a evaluar). En todo momento, se garantizó la confidencialidad de los instrumentos.

Una vez moduladas, las cajas con los instrumentos fueron trasladadas en transportes que aseguraron su arribo en adecuadas condiciones hacia las capitales de las regiones. Luego, desde la capital, estas cajas se trasladaron a las sedes provinciales. Este traslado se realizó cumpliendo un cronograma y fue monitoreado por personal del INEI y del Minedu. Una vez que las cajas llegaron a cada sede provincial, fueron guardadas en un ambiente privado y seguro, que debía permanecer cerrado bajo llave y con resguardo policial.

En paralelo a la impresión, el operador logístico conformó la RA de la siguiente manera:

- 29 coordinadores regionales y 12 coordinadores UGEL para Lima, Callao y Arequipa: responsables de organizar y ejecutar el operativo en su región / UGEL;
- 195 supervisores provinciales y 45 supervisores distritales para Lima y Callao: responsables de organizar y ejecutar el operativo en la provincia asignada, quienes también tienen bajo su responsabilidad la capacitación de los asistentes de supervisor;
- 1 589 asistentes de supervisor para los operativos de ECE y MC de segundo grado y 114 asistentes de supervisor para cuarto grado EIB: responsables de organizar y ejecutar la selección y capacitación de los aplicadores. Estos también son responsables de supervisar la aplicación y de realizar el contacto con las IE a evaluar;

- 281 asistentes administrativos: encargados de la labor administrativa y logística en la sede provincial;
- 40 301 aplicadores para los operativos de ECE y MC<sup>4</sup> de segundo grado y 1 775 aplicadores para cuarto grado EIB: responsables directos de la adecuada aplicación de los instrumentos en campo.

Para cada uno de los niveles de la RA, a excepción de los asistentes administrativos, el Minedu entregó Manuales de Aplicación referidos al operativo de campo. En ese sentido, se elaboraron siete manuales para el personal de la RA; en estos se detallaron los procesos de aplicación y las funciones que le correspondía a cada nivel:

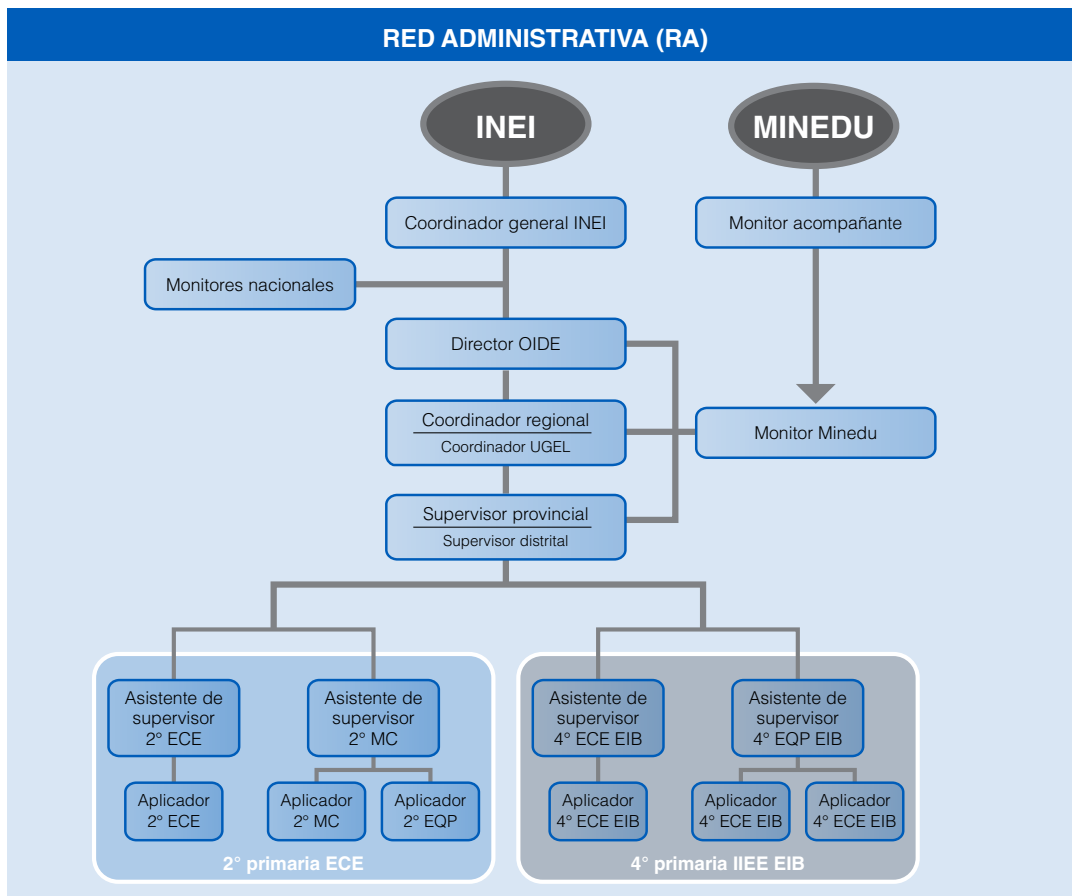
- *Manual para coordinadores y supervisores*
- *Manual para asistentes de supervisor de segundo ECE*
- *Manual para asistentes de supervisor de segundo MC*
- *Manual para asistentes de supervisor de cuarto EIB ECE*
- *Manual para aplicadores de segundo ECE*
- *Manual para aplicadores de segundo MC*
- *Manual para aplicadores de cuarto EIB ECE*

Por parte del Minedu, se implementó un monitoreo de campo que estuvo a cargo de 45 monitores regionales y provinciales. Estos monitores tuvieron como responsabilidades:

- coordinar con las instancias regionales, locales (Gobierno Regional, Dirección Regional de Educación, Unidad de Gestión Educativa Local) y RA del operador logístico para asegurar la difusión y llevar un proceso adecuado de implementación de la ECE;
- capacitar a los coordinadores regionales, supervisores provinciales y asistentes de supervisor para asegurar la incorporación de procedimientos de estandarización de aplicación de la ECE;
- monitorear la convocatoria, selección y capacitación de aplicadores;
- monitorear el desarrollo de los procesos estipulados en los manuales y en el *Plan de Aseguramiento de la Calidad y la Confidencialidad*;
- informar al Minedu sobre la implementación de la ECE en sus diferentes fases, así como cualquier imprevisto que pudiera suscitarse y que pudiera poner en riesgo el operativo.

---

<sup>4</sup> Para el operativo MC, se contó con un aplicador líder por cada IE. Este aplicador líder fue el responsable de coordinar la aplicación en la IE, lo que garantizó que los procedimientos se cumplieran según lo establecido en los manuales de aplicación.



**Figura 2. Red administrativa implementada en la ECE 2013**

#### 4.2 PROCESO DE SELECCIÓN Y CAPACITACIÓN DE APLICADORES

La convocatoria de aplicadores se realizó a través del portal web del INEI. Este sistema web permitió la inscripción de postulantes que cumplieran con el siguiente perfil:

- ✓ formación según el siguiente orden de prioridad:
  - docente con experiencia en aula, sin carga docente o administrativa en el sector estatal en la actualidad.
  - egresado universitario o técnico no docente con experiencia en cualquiera de las siguientes áreas:
    - proyectos de formación, de desarrollo dirigido a la infancia; o
    - desarrollo de proyectos artístico-educativos o de animación sociocultural dirigido a niñas y niños;
  - estudiantes de los dos últimos años de Educación, Psicología, Trabajo Social, Comunicación Social, Sociología, Antropología o Filosofía, con alguna de las siguientes experiencias: en el aula, en proyectos de formación o en actividades de proyección social con niñas y niños.
- ✓ habilidades para el buen manejo de grupos de estudiantes y de las relaciones personales con diversas personas;
- ✓ deseable experiencia en la aplicación de instrumentos estandarizados de evaluación educativa;
- ✓ buen nivel de lectura comprensiva;

- ✓ adecuada dicción;
- ✓ disposición a tiempo completo durante las fechas de capacitación y aplicación y para viajar al lugar que se le designe;
- ✓ obligatoriedad de asistir a las sesiones de capacitación;
- ✓ edad entre 20 y 50 años, salvo excepciones que deberán ser aprobadas por el Minedu;
- ✓ disponibilidad para desplazarse al lugar que se les asigne;
- ✓ deseable que disponga de teléfono móvil.

Los postulantes inscritos en el portal *web* entregaron sus hojas de vida documentada en las oficinas de las sedes provinciales. La revisión de la documentación estuvo a cargo de los asistentes de supervisor. Al finalizar esta revisión, se convocó a quienes cumplían con el perfil descrito para rendir una prueba, pero solo quienes obtuvieron nota aprobatoria fueron convocados a la capacitación de aplicadores.

Con el objetivo de seleccionar a los postulantes con mejor desempeño, se convocó a 8% de candidatos adicionales al número requerido para cada grado. Ello también permitió contar con personal suplente. Esta capacitación tuvo una duración de dos días y las aulas estuvieron conformados por un máximo de treinta candidatos.

Durante la capacitación, se instruyó a los candidatos en la aplicación estandarizada de instrumentos, para ello se utilizó el *Manual del aplicador de la ECE* y los materiales elaborados para esa finalidad: fichas, papelógrafos, guiones de aplicación, etc. Los asistentes de supervisor llevaron a cabo un detallado diseño de capacitación que permitió que en todo el país los candidatos recibieran la misma información, en el que se incluía actividades de lectura individual y grupal, actividades sobre manejo de procedimientos y documentos de aplicación, así como su simulación; también detallaba la forma de calificación que debía utilizarse para evaluar el desempeño de los candidatos en cada uno de los ejercicios propuestos. 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.

Los aplicadores fueron ordenados según los resultados de la evaluación y designados a los operativos según la siguiente prioridad:

1. aplicadores líderes MC,
2. aplicadores EQP,
3. aplicadores MC,
4. aplicadores ECE designados a instituciones educativas grandes, problemáticas o lejanas,
5. aplicadores ECE designados a instituciones educativas pequeñas.

Los candidatos con mejor desempeño fueron designados a MC y EQP, y fueron citados a medio día más de capacitación para instruirlos sobre los procedimientos propios de estos operativos.

### 4.3 PROCESO DE RECOLECCIÓN DE DATOS

Los aplicadores de IE urbanas y de fácil acceso recogieron el mismo día de la aplicación los materiales de acuerdo a un cronograma preparado por el asistente de supervisor, que les permitió, en la mayoría de los casos, llegar a la IE asignada antes del inicio de la jornada escolar. Este procedimiento se repitió durante todos los días de aplicación, dado que solo se entregaron los instrumentos necesarios para cada día.

Los aplicadores asignados a las IE rurales recibieron los materiales e instrumentos para todos los días de aplicación antes de iniciar el traslado hacia la IE.

Como se señaló al inicio de este capítulo, los cuadernillos de segundo grado fueron aplicados los días 12 y 13 de noviembre, y los cuadernillos de cuarto grado EIB, el 12 de noviembre. En una sección de las IIEE en la que se aplicó el operativo EQP, se evaluó dos días adicionales: 14 y 15 de noviembre. Cada día de aplicación se administró dos cuadernillos, tal como se muestran en las tablas que aparecen a continuación:

**Tabla 6** Orden y duración de las sesiones de aplicación, segundo grado

Grado y operativo	Primer día	Segundo día
Segundo grado ECE y MC	Comprensión lectora (45 minutos)	Matemática (40 minutos)
	Descanso (30 minutos)	Descanso (30 minutos)
	Matemática (40 minutos)	Comprensión lectora (45 minutos)

**Tabla 7** Orden y duración de las sesiones de aplicación, cuarto grado

Grado y operativo	Primer día
Cuarto grado EIB	Comprensión lectora en castellano como segunda lengua (45 minutos)
	Descanso (30 minutos)
	Comprensión lectora en castellano como segunda lengua (40 minutos)

Los aplicadores administraron los cuadernillos siguiendo la secuencia y los tiempos establecidos, además de seguir los guiones de aplicación. Luego del trabajo en aula, los aplicadores se dirigieron a un ambiente privado para realizar la verificación de los cuadernillos y el traslado de las respuestas a la ficha óptica.

Al finalizar el último día de aplicación, el director o personal responsable de la IIEE dio la conformidad de la aplicación en una Constancia de Aplicación para cada sección evaluada y verificó el correcto traslado de las respuestas a la ficha óptica. En el caso de las IIEE ECE, el aplicador entregó al director y al docente de segundo y cuarto grados EIB los cuadernillos aplicados y no aplicados. Todos los cuadernillos de MC, aplicados y no aplicados, regresaron a Lima para su respectiva verificación.

Además, para los casos en que no se concretó la aplicación, se contó con una Constancia de No Aplicación. Dicha constancia detallaba el motivo de no aplicación los que, en los casos pertinentes, deberían ser respaldados por la documentación respectiva.

Al finalizar el trabajo en la IIEE, el aplicador se dirigió a la sede provincial para entregar todos los instrumentos. El asistente de supervisor verificó el material devuelto teniendo en cuenta la cantidad y calidad de la información recogida.

Luego de este proceso, los asistentes de supervisor y supervisores clasificaron y modularon los instrumentos. Los materiales pertenecientes a IIEE de los operativos MC y EQP regresaron en las mismas cajas en las que fueron enviados, mientras que las fichas ópticas del operativo ECE fueron embaladas según la sede provincial a la que pertenecían. Al finalizar este proceso se enviaron las cajas a la empresa responsable de la captura de la información.



## 5. Propiedades psicométricas de las pruebas aplicadas

Se puede considerar que el objetivo fundamental de la medición en Psicología y Educación es describir alguna característica de las personas como un puntaje numérico. En esta perspectiva, la Psicometría puede ser definida como aquella disciplina que estudia la medición de las características psicológicas, como las habilidades, aptitudes, logro académico, rasgos de personalidad y conocimientos (Everitt y Skronidal, 2010). Por su parte, Muñiz (1996) define a la Psicometría como el conjunto de métodos, técnicas y teorías implicadas en la medición de las variables psicológicas. Lo que constituye lo específico de la Psicometría sería su énfasis y especialización en aquellas propiedades métricas exigibles a las mediciones psicológicas independientemente del campo específico de aplicación y de los instrumentos utilizados, como pueden ser aquellos aplicados en la medición de logros educativos.

En este contexto, aparecen las teorías psicométricas, que implican básicamente la aplicación de modelos matemáticos y estadísticos para su formulación, tratando de justificar los procesos subyacentes a la medición de variables que no pueden ser observadas directamente, como lo constituyen los distintos rasgos psicológicos.

En la actualidad, hay básicamente dos enfoques psicométricos fundamentales (Suen, 1990; Muñiz 1999). Estas son de acuerdo con Suen (1990) la teoría del muestreo aleatorio, que incluye a la Teoría Clásica de los Test (TCT) y a la Teoría de la Generalizabilidad; y la Teoría de Respuesta al Ítem.

La teoría del muestreo aleatorio ve la conexión del puntaje observado con el puntaje verdadero como un problema de generalización desde una muestra a una población más grande o universo. Este enfoque consta de dos modelos. El primero de ellos es la Teoría Clásica de los Test que plantea una relación de tipo lineal entre el puntaje observado, el puntaje verdadero y el error en la medición. El segundo modelo es la Teoría de la Generalizabilidad, propuesta por Cronbach y colaboradores (Suen, 1990). Muñiz (1999) afirma que se puede considerar a esta teoría como una extensión del modelo clásico que utiliza el análisis de varianza para analizar las fuentes de error de un modo sistemático y desagregado.

La teoría de respuesta al ítem sugiere que sí se puede determinar cómo cada ítem en un test opera con una persona, podemos estimar el puntaje verdadero de cada persona de manera directa. Suen (1990) afirma que dentro de esta aproximación hay tres modelos fundamentales a saber: el Modelo Logístico de 2 Parámetros y el Modelo Logístico de 3 Parámetros, ambos de Birnbaum; y el Modelo de 1 Parámetro de Lord. Los tres tratan de explicar cómo se relaciona la habilidad de una persona con la probabilidad de responder correctamente a un ítem, considerando diferentes características del ítem, como su dificultad, discriminación y pseudoadivinación (Hambleton, Swaminathan y Rogers, 1991).

Las pruebas aplicadas en la ECE son analizadas mediante el modelo Rasch para ítems dicotómicos, que se expone en el acápite que aparece a continuación.

### 5.1 EL MODELO RASCH PARA ÍTEMS DICOTÓMICOS

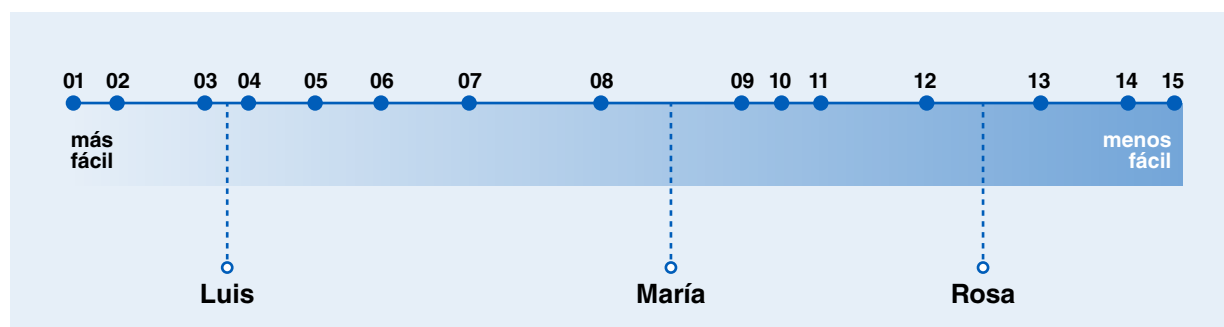
El análisis psicométrico de las pruebas aplicadas en las Evaluaciones Censales de Estudiantes (ECE) se basó en el modelo Rasch para ítems dicotómicos (Bond y Fox, 2001; Wright y Stone, 1998). Este modelo se centra en el análisis de cada ítem, concretamente de la interacción entre una persona y un ítem. Establecen 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 ( $\theta$ ) y la medida del ítem utilizado en términos de su dificultad ( $\beta$ )<sup>5</sup>. Por este motivo, se les denomina usualmente modelos de un parámetro (Hambleton, Swaminathan y Rogers, 1991). Además, es importante señalar que 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.

<sup>5</sup> En algunos textos, se utiliza la letra B para referirse a la habilidad de las personas y la letra D para referirse a la dificultad de los ítems.

George Rasch, matemático danés, se dio cuenta que los resultados de la interacción entre personas e ítems no pueden estar totalmente predeterminados, sino que implica siempre un elemento de impredecibilidad (Wright y Linacre, 1989). Esto conlleva al requerimiento que, en términos probabilísticos, mientras mayor sea la habilidad de una persona, mayor sea su probabilidad de acertar 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. Pensar con probabilidades implica un salto de lo observable y fijo a lo relativo y probable (Ingebo, 1997).

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) nos recuerdan que la existencia de una métrica común permite combinar la habilidad de la persona y la dificultad del ítem para predecir su desempeño en un ítem cualquiera e identificar respuestas inesperadas. La idea central del análisis Rasch es poder construir una escala conformada por los ítems ordenados según su dificultad. Ello implica que a mayor habilidad, la persona tendrá una mayor probabilidad de acertar a los ítems y, por lo tanto, 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 (número de ítems correctos) que posee, este será solo un insumo a partir del cual se construirá la medida Rasch.

Como ejemplo, supongamos que se ha construido una prueba con quince ítems que se ajustan a un modelo Rasch y se les ha ordenado según su dificultad, del ítem más fácil (01) al más difícil (15):



**Figura 3. Recta de ubicación de quince ítems según su dificultad y tres personas según su habilidad**

Si sabemos que Luis tiene una habilidad mayor que la dificultad del ítem 03, pero menor que las del ítem 04, lo más probable es que haya acertado al ítem 03 y todos los más fáciles (01 y 02), y haya fallado al ítem 04 y todos los más difíciles (05 al 15). Por su parte, María tiene una habilidad mayor que la dificultad del ítem 08, pero menor que la del ítem 09; por lo tanto, lo más probable es que haya acertado al ítem 08 y todos los más fáciles (01 al 07), y haya fallado al ítem 09 y todos los más difíciles (10 al 15). Finalmente, Rosa probablemente habrá acertado al ítem 12 y todos los más fáciles, y habrá fallado el 13 y los más difíciles. Nótese que hemos dicho que es probable que haya acertado todos los más fáciles y fallado los más difíciles. No estamos afirmando que en la realidad se encontrará este tipo de patrones de fallos y aciertos. Lo usual es, por ejemplo, que María haya acertado los ítems 01 a 05, haya fallado el 06, acertado los 07 y 08, fallado el 09, acertado el 10 y fallado todos los demás.

Como se señaló al principio de este acápite, se ha trabajado con el modelo Rasch para ítems dicotómicos. Un ítem dicotómico tiene una sola respuesta correcta; por lo tanto, se puede acertar y recibir un punto ( $X=1$ ) o fallar y no recibir ningún puntaje ( $X=0$ ).

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

$$1. \quad \theta > \beta; p(X=1 | \theta, \beta) \in ]0,5 ; 1,0[ \quad (5.1)$$

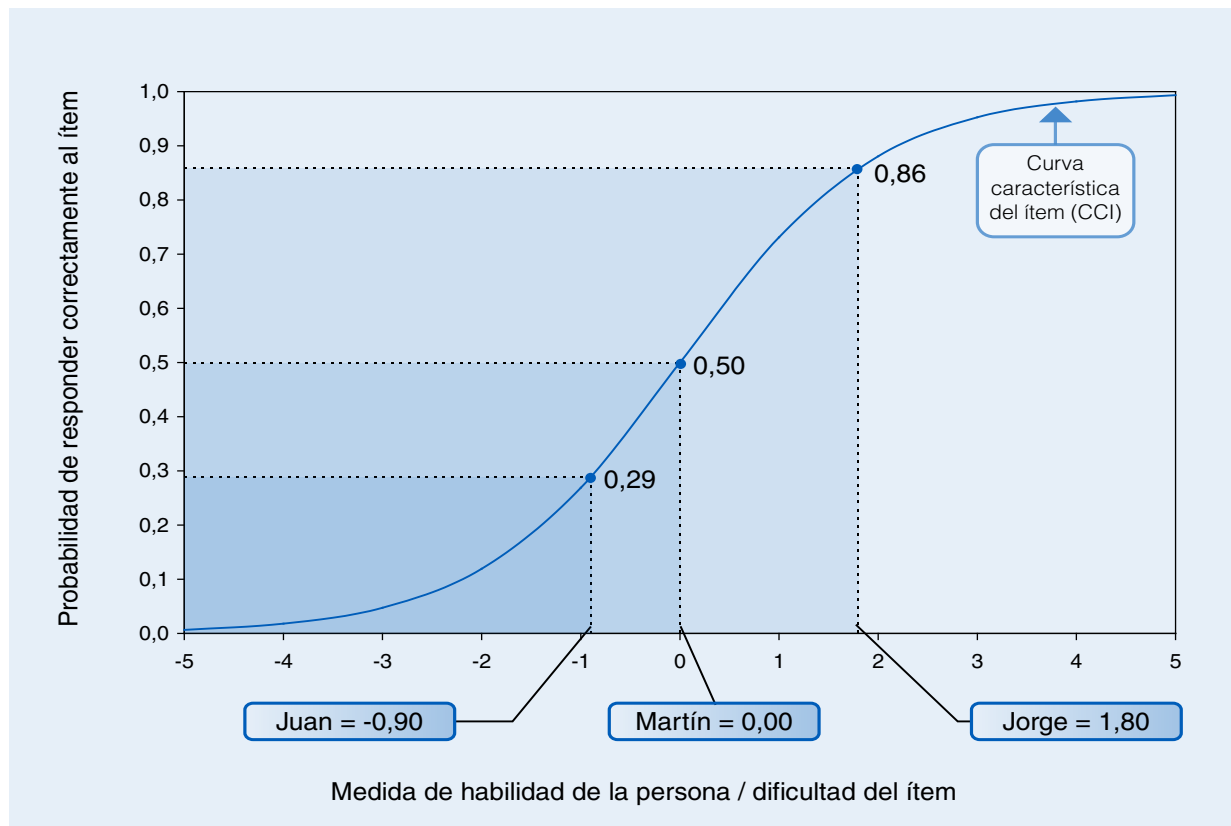
$$2. \quad \theta < \beta; p(X=1 | \theta, \beta) \in [0,0 ; 0,5[ \quad (5.2)$$

$$3. \quad \theta = \beta; p(X=1 | \theta, \beta) = 0,5 \quad (5.3)$$

El primer caso nos dice que si la habilidad de la persona es mayor que la dificultad del ítem, la probabilidad de responder correctamente a dicho ítem es mayor que 0,5 (50%). La segunda situación indica que si la habilidad de la persona es menor que la dificultad del ítem, la probabilidad de responder correctamente a dicho ítem es menor que 0,5 (50%). Finalmente, si la habilidad de la persona es igual que la dificultad del ítem, la probabilidad de responder correctamente a dicho ítem es igual a 0,5 (50%). Como señala Ingebo (1990), de esta manera se puede comprobar empíricamente la teoría que los estudiantes con mayores conocimientos tienen una mayor probabilidad de responder correctamente a una pregunta, frente a los estudiantes con menor conocimiento. Matemáticamente, la CCI se grafica con la siguiente función:

$$P(X_{ni} = 1 | \theta_n, \beta_i) = \frac{e^{\theta_n - \beta_i}}{1 + e^{\theta_n - \beta_i}} \quad (5.4)$$

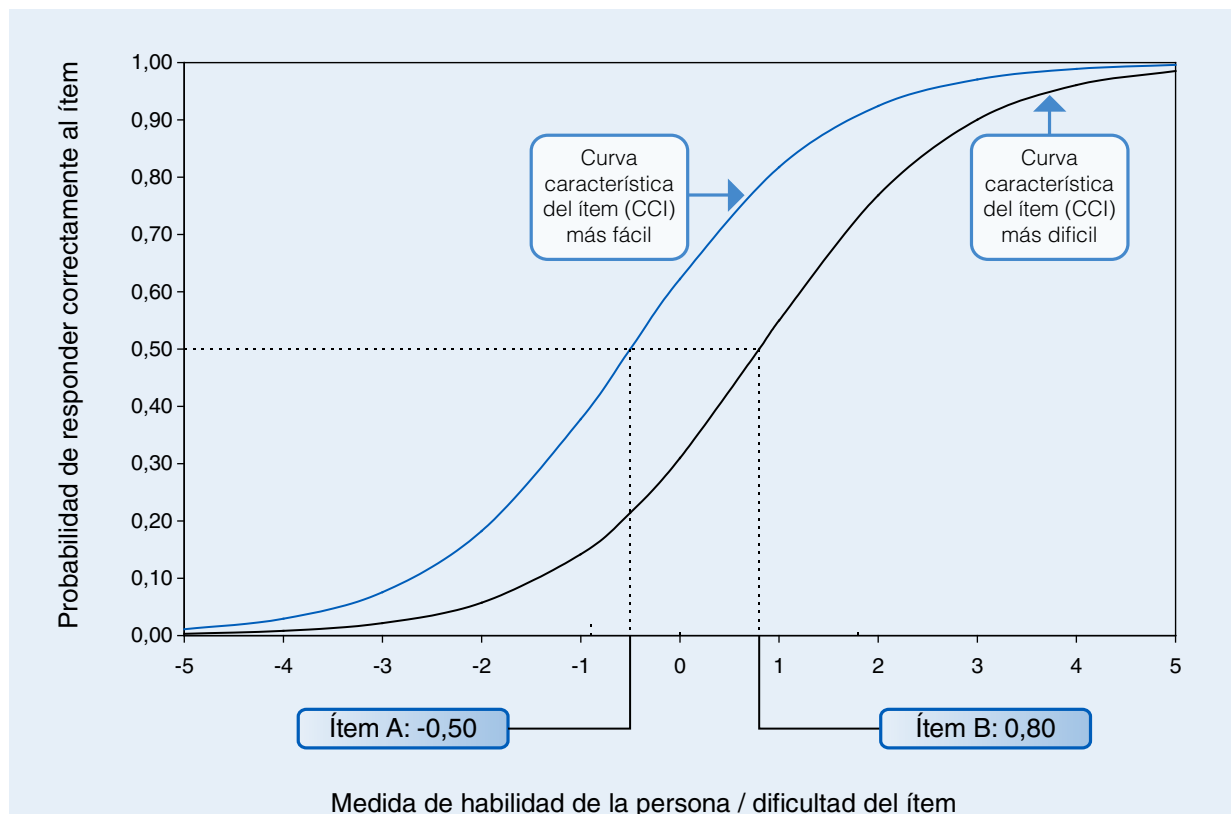
Donde  $\theta$  representa el rasgo latente de cada persona  $n$ , y  $\beta$  la medida de dificultad de cada ítem  $i$ . Esta relación entre la habilidad de una persona y la dificultad de un ítem se presenta en el siguiente gráfico:



**Figura 4. Curva característica del ítem y probabilidades de tres personas de responder correctamente dicho ítem**

En la figura anterior, se ve que, al enfrentarse a este ítem en particular, Juan, cuya habilidad es de -0,90, tiene una probabilidad de 0,29 de acertar a este ítem, 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 0,89 de acertar a este ítem; por lo tanto, es más probable que lo acierte y reciba un punto. Finalmente, Martín tiene una habilidad igual a la dificultad del ítem, por eso se afirma que tiene iguales posibilidades de acertar o de fallar el ítem.

Al comparar dos o más curvas características de ítems, se puede decir que, mientras más a la derecha se encuentra una curva, más difícil es el ítem. Además, se expresa numéricamente la dificultad de un ítem, como aquel valor de la habilidad que posee una probabilidad de 50% de acertar a dicho ítem. En el caso que se presenta en la figura que aparece a continuación, el ítem más fácil (A) tiene una dificultad de -0,50, y el más difícil (B), de 0,80:



**Figura 5. Curva característica de dos ítems comparando la dificultad de los mismos**

Como se señaló con anterioridad,  $\theta$  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. Como ello no es posible, en términos empíricos lo que se tiene es una estimación de dicha habilidad, representada como  $\hat{\theta}$ . Además,  $\beta$  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 de la siguiente manera:

$$P_{ni} = P(X_{ni} = 1 | \hat{\theta}_n, b_i) = \frac{e^{\hat{\theta}_n - b_i}}{1 + e^{\hat{\theta}_n - b_i}} \quad (5.5)$$

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

$$e. e. (b_i) = \frac{1}{\sqrt{\sum_{n=1}^{I-1} N_n P_{ni} (1 - P_{ni})}} \quad (5.6)$$

Un concepto muy importante en los modelos Rasch es el de los residuos o residuales. En estadística, los residuales son definidos como la diferencia entre los valores observados y los valores esperados bajo un modelo específico (Miles, 2005). Tomando en cuenta esto, el modelo Rasch usa los residuos de respuesta individual. El residuo bruto para una persona concreta en un ítem específico ( $R_{ni}$ ) se calcula de la siguiente manera (Wilson, 2005):

$$R_{ni} = X_{ni} - P_{ni} \quad (5.7)$$

Donde  $X_{ni}$  es la respuesta observada para la persona  $n$  en el ítem  $i$ , y  $P_{ni}$  es la respuesta esperada según el modelo para esa misma persona en ese mismo ítem. Ya que los parámetros del ítem y la persona son

desconocidos, en la práctica se usa  $\hat{P}_{ni}$  en lugar de  $P_{ni}$  para el cálculo de los residuos (Christensen y Kreiner, 2013). Además, los residuos estandarizados se calculan según la fórmula que se presenta a continuación (Wright y Stone, 1999):

$$Z_{ni} = \frac{X_{ni} - P_{ni}}{\sqrt{P_{ni}(1 - P_{ni})}} = \frac{R_{ni}}{\sqrt{VAR(P_{ni})}} \quad (5.8)$$

Finalmente, los residuales estandarizados cuadráticos quedan definidos de la siguiente manera:

$$Z_{ni}^2 = \frac{R_{ni}^2}{VAR(P_{ni})} \quad (5.9)$$

El concepto de residuo es muy importante, pues como se verá en los acápites que aparecerán posteriormente en este reporte técnico, se utilizan para evaluar el ajuste de los datos al modelo, lo que facilita calcular diversas medidas que permiten tomar decisiones referidas a mantener o descartar un ítem, o considerar si los ítems se comportan de manera unidimensional o no.

## 5.2 EL MODELO RASCH FRENTE A OTROS MODELOS PSICOMÉTRICOS

Como se señaló al principio, existen diversos modelos de análisis psicométrico. ¿Por qué utilizar el modelo Rasch y no otros modelos? Se tratará de responder esta pregunta exponiendo de manera concisa las diferencias y ventajas del modelo Rasch frente a otros modelos.

### 5.2.1 EL MODELO RASCH FRENTE A LA TEORÍA CLÁSICA DE LOS TEST (TCT)

#### 5.2.1.1 INDEPENDENCIA DE LAS PERSONA Y DEL TEST

En la TCT, la dificultad de los ítems depende de la habilidad de las personas en las cuales fueron calculados y las estimaciones de las habilidades de las personas dependen de los ítems (prueba) con los cuales fueron evaluados. Es decir, hay una dependencia del test y dependencia de las personas (Muñiz, 1997).

En el modelo Rasch, no ocurre esta dependencia, pues se caracteriza por una propiedad conocida como *objetividad*. La objetividad implica que las medidas sean independientes de los agentes (ítems) utilizados para realizar la medición y que los agentes de medición sean independientes de las personas utilizadas para medirlos (Shaw, 1991). Esta propiedad se puede demostrar matemáticamente, tal y como se aprecia en el artículo de Wright y Linacre (1987).

#### 5.2.1.2 MEDICIÓN DE INTERVALO

En la Teoría Clásica de los Test, se considera que el puntaje directo es una expresión directa del rasgo o atributo latente que se pretende medir con un test, con el cual mantiene una relación monótonica lineal (Muñiz, 1997). Sin embargo, uno puede preguntarse si la cantidad de rasgo latente que se necesita para pasar de un puntaje directo de 2 a otro de 4 no es la misma que se necesita para pasar de 15 a 17 puntos, pues eso dependerá de la dificultad de los ítems. Por lo tanto, las puntuaciones directas no constituyen una verdadera escala de intervalo, tal y como la define Stevens (1951). A pesar de ello, son utilizados en múltiples procedimientos estadísticos, que tienen entre sus supuestos que la variable analizada está medida por lo menos en un nivel de intervalo.

Si bien los puntajes son esenciales para construir medidas, no son medidas por sí mismos (Wright y Linacre, 1989), mientras que el modelo Rasch sí logra construir una escala de intervalo gracias a sus distintas propiedades matemáticas (Wright y Linacre, 1987). El *logit* constituye la unidad básica del modelo Rasch, que establece un aumento a razón de 2,718 en el cociente que se obtiene al dividir la probabilidad de responder

correctamente a un ítem entre la probabilidad de fallarlo (Linacre y Wrigth, 1989).

## **5.2.2 EL MODELO RASCH FRENTE A LA TEORÍA DE RESPUESTA AL ÍTEM (IRT)**

### **5.2.2.1 PRESCRIPCIÓN FRENTE A DESCRIPCIÓN**

Algunos afirman que el modelo Rasch es una versión simplificada de otros modelos IRT, como los de dos y tres parámetros (Hambleton, Swaminathan y Rogers, 1991). El modelo Rasch aparecería cuando el tercer parámetro es fijado en 0 y el segundo parámetro, en 1. Pero el hecho que el modelo Rasch se pueda derivar matemáticamente de otros modelos con más parámetros, no implica que este sea una versión simplificada de los mismos. La racionalidad del modelo Rasch tiene que ver con la teoría de la medición. No interesa ver que tan bien se ajusta el modelo Rasch a los datos, lo que interesa es ver si los datos se ajustan al modelo Rasch. Esta es la diferencia entre un modelo descriptivo (como los modelos TRI), que trata de explicar la mayor cantidad posible de varianza, y un modelo prescriptivo, que propone una teoría sobre qué es medir, y trata de verificar si los datos se ajustan bien al modelo, con lo cual los datos serán una buena medida (Shaw, 1991).

### **5.2.2.2 EL SEGUNDO PARÁMETRO O DISCRIMINACIÓN**

En el modelo IRT de dos parámetros el segundo parámetro denominado discriminación, se operacionaliza mediante la pendiente de la curva característica del ítem (Hambleton, Swaminathan y Rogers, 1991). Esto genera que al haber varios ítems en una prueba, sus CCI se crucen. Por lo tanto, el ordenamiento según la dificultad de los ítems dependerá de la medida concreta de habilidad de una persona. Con ello se introducen otras dimensiones en el modelo, yendo en contra del supuesto de unidimensionalidad, que es esencial para poder construir una medida (Bond y Fox, 2001).

Perkins y Engelhard (2009) señalan que el uso del segundo parámetro va en contra de dos requerimientos importantes de un sistema de medición: las medidas de las personas deben ser independientes de los ítems utilizados, una persona con mayor medida debe tener siempre una mayor probabilidad de responder correctamente a un ítem que una persona con una menor medida de habilidad. Estos dos requerimientos solo se cumplen si las curvas características de los ítems no se cruzan, como pasa con el modelo Rasch.

Un problema adicional con el modelo de dos parámetros es que a menos que se pongan limitaciones explícitas, no se logra la convergencia de los parámetros (Shaw, 1991). El segundo parámetro solo converge con algunas limitaciones que se imponen artificialmente en los algoritmos para el cálculo de dicho parámetro (Kelley, Ebel y Linacre, 2002).

### **5.2.2.3 EL TERCER PARÁMETRO O SEUDOADIVINACIÓN**

El problema es que el uso de la asíntota inferior como parámetro de adivinación requiere asumir que esta es una característica del ítem, que se mantiene fija en todas las personas (Hambleton, Swaminathan y Rogers, 1991). Es decir, que todas las personas que se enfrentan al ítem tienen la misma tendencia a adivinar. Este parámetro depende mucho de las personas que han sido utilizadas para calibrar los ítems, cuya tendencia a adivinar puede variar mucho de muestra a muestra. Por lo tanto, el tercer parámetro no es independiente de las personas utilizadas para calibrar el test (Wright, 1988).

En el modelo Rasch, se puede observar qué personas tienen un mal ajuste al modelo porque han adivinado un ítem difícil, teniendo una baja habilidad. Con ello, la tendencia a adivinar se convierte en una característica de la persona y no en una característica del test (Wright y Stone, 1999).

### **5.2.2.4 EL PROBLEMA DE LA SUFICIENCIA**

En estadística, la suficiencia implica que un conjunto de datos tiene toda la información disponible. En el modelo Rasch, esto implica que el puntaje directo contiene toda la información sobre la habilidad de la persona evaluada. Esto no ocurre con los modelos de dos o tres parámetros, en los cuales el patrón de respuestas tiene toda la información disponible (Wright, 1989). ¿Cómo podría explicarle a un estudiante o a un padre de familia

que, a pesar de haber contestado la misma cantidad de respuestas correctas en una prueba, dos estudiantes se encuentran en distintos niveles de desempeño por haber respondido diferentes preguntas? Lo anterior no ocurre con los modelos Rasch.

### 5.3 PROGRAMA INFORMÁTICO UTILIZADO PARA EL ANÁLISIS RASCH

Los análisis se han realizado mediante el programa informático Winsteps 3.80.1 (Linacre, 2013), que sirve para ajustar los diferentes modelos Rasch a un conjunto de respuestas, ya sean dicotómicas o politómicas. Este programa permite, entre otras cosas, trabajar con respuestas de opción múltiple (tipo A, B, C, D), introduciendo la secuencia de claves correctas.

La calibración de los ítems en Winsteps comienza con un estimado central para cada calibración de persona e ítem. Estos estimados iniciales se producen empleando el método *Normal Approximation Algorithm* (PROX) para alcanzar una convergencia aproximada al patrón de datos observado. Este procedimiento aprovecha la forma similar que tienen la función logística y la curva normal. Modela juntos a las personas e ítems para que ambos estén distribuidos normalmente. La variante del PROX implementado en Winsteps permite valores perdidos y su ecuación para las estimaciones es la siguiente:

$$B_n = \mu_n + \sqrt{(1 + \sigma_n^2/2.9)} \times \ln \{R_n/(N_n - R_n)\} \quad (5.10)$$

Donde  $R_n$  es el puntaje total logrado por la persona  $n$  en  $N_n$  ítems,  $\mu_n$  y  $\sigma_n$  resumen la distribución del *logit* de la dificultad del ítem encontrada por la persona  $n$ . En Winsteps, las iteraciones PROX cesan cuando la varianza de los ítems no incrementa substancialmente de una iteración a otra.

Una vez que se tienen los valores iniciales con el método PROX, se aplica el procedimiento *Joint Maximum Likelihood Estimation* (JLME) de forma iterada para obtener mayor exactitud en las estimaciones de los parámetros de las personas e ítems. Adicionalmente, con el método JLME se obtienen los errores estándar de los estadísticos estimados y medidas de ajuste de los datos. La implementación de JLME se basa en los métodos de estimación de máxima verosimilitud incondicional y máxima verosimilitud conjunta para llegar a los estimados finales de los parámetros de personas e ítems. El proceso de estimación es iterativo y el algoritmo se presenta de la siguiente manera:

$$L_u = \frac{\exp(\sum_n \theta_n r_n) \exp(-\sum_i \beta_i s_i)}{\prod_n \prod_i (1 + \exp(\theta_n - \beta_i))}$$

Donde los estadísticos suficientes son:  $rn = \sum_i x_{ni}$  para  $\theta_n$  y  $si = \sum_n x_{ni}$  para  $\beta_i$ .

### 5.4 ANÁLISIS DE ÍTEMS

En términos generales, el análisis de ítems puede seguir dos perspectivas (Prieto y Delgado, 1996):

- Cualitativa: se centra en el análisis de los aspectos formales del ítem, como su redacción y ortografía. Además, considera aspectos de contenido, como la pertinencia del ítem para el grupo de personas evaluadas y la posible presencia de sesgo.
- Cuantitativa: aplica un conjunto de procedimientos estadísticos que permiten cuantificar distintas características psicométricas de los ítems, como su dificultad y discriminación, entre otras.

El análisis cualitativo fue realizado por expertos con experiencia en la construcción de instrumentos de medición del logro en las distintas áreas evaluadas de la ECE; mientras que los análisis cuantitativos son presentados en este reporte técnico.

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 ver al 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, 2007).

Como lo señala Wilson (2005), el ajuste de los ítems a un modelo Rasch se basa en dos estadísticos, el Outfit e Infit, cuyo cálculo se basa en los residuos cuadráticos estandarizados del modelo (. Basándose en las propuestas de diversos autores (Bond y Fox, 2007; Linacre y Wright, 1994; Wilson 2005), se expone brevemente el significado y 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 es como sigue:

$$Outfit_i = \frac{1}{N} \sum_{n=1}^N Z_{ni}^2 \quad (5.11)$$

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:

$$Infit_i = \frac{\sum_{n=1}^N R_{ni}^2}{\sum_{n=1}^N VAR(P_{ni})} \quad (5.12)$$

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:

$$Infit_i = \sum_{n=1}^N \omega_{ni} Z_{ni}^2 \quad (5.13)$$

Donde los residuales están ponderados por sus varianzas individuales:

$$\omega_{ni} = \frac{VAR(P_{ni})}{\sum_{n=1}^N VAR(P_{ni})} \quad (5.14)$$

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, también 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, 2005):

$$ptme = \frac{\sum_{n=1}^N [(X_n - \sum_{m=1}^N X_m/N)(\hat{\theta}_n - \sum_{m=1}^N \hat{\theta}_m/N)]}{\sqrt{\sum_{n=1}^N [(X_n - \sum_{m=1}^N X_m/N)^2] \sum_{n=1}^N [(\hat{\theta}_n - \sum_{m=1}^N \hat{\theta}_m/N)^2]}} \quad (5.15)$$

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

Siguiendo las recomendaciones de Linacre (2005), se espera que los ítems con un ajuste adecuado al modelo Rasch tengan valores Infit y Outfit entre 0,50 y 1,50, pero dentro de este grupo se prefiere 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 2005).

Se presenta en la tabla que aparece a continuación los diferentes indicadores de ajuste calculados para los ítems que conforman las distintas pruebas aplicadas como parte de la ECE. Además, se presenta 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 fluctúan entre 0 y 1, y los valores más bajos representan ítems más difíciles. Este índice para un ítem 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 ( $N_i$ ).

$$p_i = \frac{A_i}{N_i} \quad (5.16)$$



Finalmente, se señala en esta tabla el nivel de desempeño al que corresponde un ítem según su medida de dificultad: nivel 1 (En Proceso), nivel 2 (Satisfactorio) o si es un ítem superior al nivel 2, lo cual se indica como nivel 3. Todos los análisis de ítems fueron realizados usando una muestra aleatoria simple de 3 000 estudiantes. Se presentan las medidas ya equiparadas según los procedimientos descritos en el acápite 5.7.

**Tabla 8** Dificultad y ajuste de los ítems al modelo Rasch, Comprensión lectora ECE

Cuadernillo	Ítem	Medida	Error	Infit	Outfit	ptme	p	Nivel
1	1	-1,237	0,071	1,05	0,90	0,36	0,92	1
1	2	-0,729	0,062	1,07	0,93	0,39	0,88	1
1	3	-0,437	0,058	0,89	0,87	0,44	0,86	1
1	4	-1,300	0,073	0,99	0,64	0,42	0,92	1
1	5	-0,278	0,056	0,92	0,82	0,48	0,84	1
1	6	-0,035	0,053	1,03	1,06	0,45	0,80	1
1	7	-0,393	0,057	0,82	0,71	0,47	0,86	1
1	8	0,418	0,049	1,03	1,12	0,41	0,76	2
1	9	0,186	0,051	0,96	0,91	0,43	0,80	2
1	10	1,852	0,043	1,16	1,26	0,39	0,54	2
1	11	0,858	0,046	1,12	1,32	0,37	0,70	2
1	12	0,455	0,048	1,06	1,25	0,38	0,76	2
1	13	1,077	0,045	1,22	1,36	0,34	0,65	2
1	14	0,906	0,045	0,97	0,95	0,47	0,70	2
1	15	1,601	0,043	0,99	0,99	0,49	0,59	2
1	16	1,158	0,044	1,11	1,19	0,41	0,64	2
1	17	0,899	0,046	0,96	0,93	0,50	0,69	2
1	18	1,104	0,045	0,98	0,99	0,48	0,66	2
1	19	1,646	0,043	1,06	1,07	0,45	0,58	2
1	20	0,086	0,052	0,95	0,89	0,45	0,81	2
1	21	2,282	0,043	1,16	1,40	0,39	0,46	3
1	22	--	--	--	--	--	--	--
1	23	2,693	0,044	1,17	1,44	0,39	0,40	3
2	1	-1,183	0,07	1,11	1,33	0,33	0,91	1
2	2	-0,267	0,055	0,97	0,86	0,45	0,83	1
2	3	-0,892	0,065	0,90	0,78	0,44	0,90	1
2	4	-1,278	0,072	0,98	0,65	0,41	0,92	1
2	5	-0,319	0,056	0,98	0,89	0,44	0,84	1
2	6	-0,557	0,059	0,92	0,70	0,48	0,86	1
2	7	0,282	0,05	1,00	0,91	0,48	0,76	2
2	8	1,468	0,043	1,09	1,12	0,42	0,61	2
2	9	1,680	0,043	1,14	1,21	0,40	0,57	2
2	10	2,236	0,043	1,11	1,23	0,43	0,48	3
2	11	0,878	0,046	1,06	1,04	0,43	0,69	2
2	12	1,494	0,043	1,09	1,14	0,42	0,60	2
2	13	1,076	0,045	0,90	0,83	0,54	0,66	2
2	14	0,655	0,047	0,97	1,02	0,47	0,73	2
2	15	0,190	0,051	0,90	0,83	0,49	0,79	2
2	16	1,087	0,045	1,11	1,14	0,40	0,66	2
2	17	1,331	0,044	0,98	0,96	0,49	0,62	2
2	18	0,633	0,047	0,99	0,96	0,46	0,73	2
2	19	1,155	0,044	0,96	0,95	0,49	0,66	2
2	20	0,770	0,046	0,89	0,78	0,52	0,72	2
2	21	0,940	0,045	1,10	1,11	0,42	0,68	2
2	22	1,353	0,044	1,04	1,03	0,45	0,62	2
2	23	1,096	0,045	1,07	1,11	0,42	0,66	2

El ítem número 22 del cuadernillo 1 de Comprensión lectora no muestra un buen ajuste al modelo Rasch, por lo cual no ha sido incluido en la estimación de las medidas de los estudiantes evaluados.

**Tabla 9** Dificultad y ajuste de los ítems al modelo Rasch, Comprensión lectora MC

Cuadernillo	Ítem	Medida	Error	Infit	Outfit	ptme	p	Nivel
1	1	-1,199	0,075	0,96	0,70	0,36	0,92	1
1	2	-0,550	0,061	0,94	0,76	0,40	0,88	1
1	3	-0,510	0,061	0,86	0,75	0,45	0,87	1
1	4	-1,123	0,073	0,82	0,47	0,42	0,92	1
1	5	-0,189	0,056	0,86	0,70	0,48	0,84	1
1	6	0,209	0,051	0,91	0,86	0,46	0,80	2
1	7	-0,476	0,060	0,84	0,69	0,47	0,87	1
1	8	0,439	0,049	1,01	1,06	0,41	0,77	2
1	9	0,231	0,051	0,95	0,87	0,43	0,80	2
1	10	1,838	0,043	1,18	1,27	0,36	0,53	2
1	11	0,798	0,046	1,12	1,30	0,36	0,70	2
1	12	0,353	0,050	1,04	1,18	0,38	0,78	2
1	13	0,983	0,045	1,12	1,18	0,37	0,68	2
1	14	0,908	0,046	0,97	0,94	0,47	0,70	2
1	15	1,527	0,043	0,97	0,95	0,50	0,59	2
1	16	1,290	0,044	1,08	1,11	0,42	0,63	2
1	17	1,000	0,045	0,94	0,89	0,49	0,68	2
1	18	1,183	0,044	0,99	0,99	0,48	0,64	2
1	19	1,718	0,043	1,06	1,09	0,43	0,56	2
1	20	0,168	0,052	0,95	0,86	0,45	0,80	2
1	21	2,347	0,043	1,14	1,42	0,37	0,43	3
1	22	--	--	--	--	--	--	--
1	23	2,549	0,044	1,13	1,33	0,37	0,39	3
2	1	-1,470	0,084	0,90	0,60	0,37	0,94	1
2	2	-0,971	0,071	0,82	0,58	0,41	0,92	1
2	3	1,689	0,043	1,22	1,36	0,33	0,55	2
2	4	-0,823	0,068	1,23	1,46	0,39	0,87	1
2	5	-0,367	0,059	0,80	0,62	0,48	0,87	1
2	6	0,980	0,046	1,07	1,09	0,41	0,68	2
2	7	-0,878	0,069	0,86	0,68	0,43	0,90	1
2	8	-0,608	0,063	0,79	0,58	0,48	0,89	1
2	9	-0,349	0,059	0,83	0,68	0,46	0,86	1
2	10	0,980	0,046	1,03	0,99	0,46	0,67	2
2	11	0,844	0,047	0,96	0,89	0,50	0,69	2
2	12	0,671	0,048	1,04	1,01	0,43	0,73	2
2	13	0,969	0,046	1,00	0,99	0,46	0,68	2
2	14	--	--	--	--	--	--	--
2	15	0,799	0,047	0,92	0,84	0,51	0,71	2
2	16	0,314	0,051	0,88	0,76	0,51	0,78	2
2	17	2,089	0,043	1,16	1,26	0,37	0,48	3
2	18	1,810	0,043	1,14	1,23	0,39	0,53	2
2	19	1,453	0,044	0,92	0,88	0,53	0,59	2
2	20	0,795	0,047	0,91	0,79	0,54	0,70	2
2	21	1,974	0,043	1,06	1,13	0,44	0,51	2
2	22	0,694	0,048	1,01	0,98	0,42	0,74	2
2	23	1,666	0,043	1,17	1,30	0,36	0,55	2

Los ítems números 22 del cuadernillo 1 y 14 del cuadernillo 2 de Comprensión lectora no muestran un buen ajuste al modelo Rasch, por lo cual no han sido incluidos en la estimación de las medidas de los estudiantes evaluados.

**Tabla 10** Dificultad y ajuste de los ítems al modelo Rasch, matemática ECE

Cuadernillo	Ítem	Medida	Error	Infit	Outfit	ptme	p	Nivel
1	1	-2,284	0,071	0,94	1,07	0,33	0,92	1
1	2	-1,432	0,058	1,01	1,24	0,38	0,85	1
1	3	--	--	--	--	--	--	--
1	4	-0,462	0,048	1,04	1,18	0,46	0,72	1
1	5	-0,993	0,052	0,99	1,32	0,42	0,79	1
1	6	-0,879	0,050	0,98	0,97	0,45	0,78	1
1	7	0,388	0,045	1,18	1,29	0,44	0,58	2
1	8	0,161	0,046	0,91	0,86	0,57	0,62	2
1	9	-0,581	0,048	0,99	0,95	0,48	0,74	1
1	10	--	--	--	--	--	--	--
1	11	-0,186	0,047	0,85	0,77	0,58	0,68	1
1	12	1,492	0,046	0,96	1,04	0,58	0,39	2
1	13	0,396	0,045	1,05	1,04	0,51	0,58	2
1	14	-0,884	0,051	1,07	1,23	0,40	0,78	1
1	15	-0,466	0,048	0,86	0,82	0,55	0,72	1
1	16	2,329	0,051	0,98	1,18	0,55	0,27	3
1	17	0,426	0,045	0,94	0,89	0,57	0,58	2
1	18	0,743	0,045	1,01	1,02	0,54	0,52	2
1	19	0,700	0,045	1,12	1,21	0,49	0,53	2
1	20	0,033	0,046	0,87	0,78	0,59	0,64	1
1	21	0,965	0,045	1,01	1,04	0,55	0,48	2
2	1	-1,516	0,059	0,98	1,13	0,39	0,85	1
2	2	--	--	--	--	--	--	--
2	3	-0,168	0,046	1,16	1,47	0,41	0,67	1
2	4	-0,885	0,050	1,01	1,12	0,43	0,78	1
2	5	-1,534	0,058	0,97	1,07	0,39	0,86	1
2	6	-1,744	0,061	0,94	0,97	0,39	0,88	1
2	7	-0,666	0,050	0,86	0,76	0,54	0,75	1
2	8	1,507	0,046	0,92	0,97	0,60	0,39	2
2	9	0,743	0,045	0,96	0,98	0,57	0,52	2
2	10	-0,680	0,049	0,85	0,77	0,54	0,75	1
2	11	-1,526	0,058	0,98	1,05	0,39	0,85	1
2	12	-0,319	0,047	0,94	0,96	0,52	0,70	1
2	13	0,711	0,045	0,90	0,87	0,60	0,53	2
2	14	-0,133	0,046	1,02	1,04	0,50	0,67	1
2	15	1,638	0,047	1,09	1,20	0,51	0,37	2
2	16	-0,139	0,047	1,04	1,15	0,48	0,67	1
2	17	0,535	0,045	1,00	0,97	0,55	0,56	2
2	18	0,571	0,045	0,92	0,87	0,59	0,55	2
2	19	1,131	0,045	0,86	0,81	0,63	0,45	2
2	20	1,251	0,046	1,27	1,39	0,43	0,43	2
2	21	0,818	0,045	1,22	1,45	0,43	0,51	2

Los ítems números 3 y 10 del cuadernillo 1 y 2 del cuadernillo 2 de Matemática no muestran un buen ajuste al modelo Rasch, por lo cual no han sido incluidos en la estimación de las medidas de los estudiantes evaluados.

**Tabla 11** Dificultad y ajuste de los ítems al modelo Rasch, matemática MC

Cuadernillo	Ítem	Medida	Error	Infit	Outfit	ptme	p	Nivel
1	1	-2,114	0,075	0,94	1,00	0,32	0,92	1
1	2	-1,274	0,057	0,99	1,14	0,38	0,84	1
1	3	--	--	--	--	--	--	--
1	4	-0,385	0,048	1,02	1,16	0,47	0,71	1
1	5	-0,817	0,051	1,00	1,42	0,42	0,78	1
1	6	-0,751	0,051	0,98	1,00	0,45	0,77	1
1	7	0,485	0,045	1,21	1,35	0,45	0,56	2
1	8	0,227	0,046	0,91	0,87	0,57	0,61	2
1	9	-0,517	0,048	0,98	0,96	0,48	0,74	1
1	10	--	--	--	--	--	--	--
1	11	-0,061	0,047	0,87	0,83	0,57	0,66	1
1	12	1,509	0,047	0,98	1,05	0,58	0,37	2
1	13	0,451	0,045	1,07	1,08	0,51	0,56	2
1	14	-0,790	0,051	1,08	1,33	0,39	0,78	1
1	15	-0,393	0,048	0,85	0,80	0,55	0,72	1
1	16	2,325	0,053	0,98	1,17	0,57	0,25	3
1	17	0,333	0,045	0,92	0,88	0,58	0,59	2
1	18	0,789	0,045	1,02	1,04	0,55	0,50	2
1	19	0,690	0,045	1,13	1,25	0,49	0,52	2
1	20	0,124	0,046	0,88	0,80	0,58	0,62	2
1	21	0,936	0,045	1,04	1,10	0,55	0,47	2
2	1	-0,862	0,053	1,07	1,61	0,38	0,79	1
2	2	--	--	--	--	--	--	--
2	3	-1,216	0,058	0,98	1,13	0,39	0,84	1
2	4	--	--	--	--	--	--	--
2	5	-1,463	0,060	0,98	1,20	0,36	0,87	1
2	6	-1,298	0,059	0,95	0,92	0,41	0,85	1
2	7	0,156	0,047	0,91	0,94	0,56	0,62	2
2	8	-0,381	0,049	0,85	0,81	0,55	0,72	1
2	9	0,721	0,046	1,06	1,14	0,53	0,52	2
2	10	-0,107	0,048	1,00	1,16	0,49	0,67	1
2	11	1,676	0,049	0,90	0,87	0,63	0,35	2
2	12	0,507	0,046	0,92	0,87	0,59	0,56	2
2	13	0,590	0,046	0,92	0,95	0,59	0,54	2
2	14	1,588	0,048	1,13	1,37	0,51	0,36	2
2	15	0,644	0,046	0,87	0,80	0,62	0,53	2
2	16	-0,632	0,050	0,94	0,91	0,48	0,76	1
2	17	0,824	0,046	1,09	1,10	0,52	0,50	2
2	18	0,022	0,047	0,83	0,83	0,59	0,65	1
2	19	0,807	0,046	1,23	1,41	0,45	0,50	2
2	20	1,207	0,047	0,91	0,91	0,62	0,43	2
2	21	0,745	0,046	1,25	1,38	0,44	0,51	2

Los ítems números 3 y 10 del cuadernillo 1, y 2 y 4 del cuadernillo 2 de Matemática no muestran un buen ajuste al modelo Rasch, por lo cual no han sido incluidos en la estimación de las medidas de los estudiantes evaluados.

**Tabla 12** Dificultad y ajuste de los ítems al modelo Rasch, castellano como segunda lengua

Cuadernillo	Ítem	Medida	Error	Infit	Outfit	ptme	p	Nivel
1	1	-1,204	0,020	0,89	0,81	0,45	0,78	1
1	2	-1,535	0,021	0,92	0,81	0,41	0,83	1
1	3	-1,756	0,023	0,90	0,71	0,41	0,85	1
1	4	-0,884	0,018	0,97	0,99	0,42	0,73	1
1	5	-0,763	0,018	0,80	0,70	0,55	0,71	1
1	6	-1,347	0,020	0,91	0,84	0,43	0,80	1
1	7	0,110	0,017	1,08	1,08	0,40	0,53	2
1	8	0,405	0,017	1,23	1,32	0,30	0,47	2
1	9	-0,756	0,018	0,86	0,74	0,51	0,70	1
1	10	1,086	0,018	1,10	1,19	0,39	0,34	3
1	11	0,679	0,017	1,20	1,31	0,32	0,41	2
1	12	-0,196	0,017	1,21	1,24	0,30	0,60	1
1	13	0,427	0,017	1,00	1,03	0,46	0,47	2
1	14	0,126	0,017	1,03	1,00	0,44	0,53	2
1	15	0,724	0,017	1,09	1,10	0,41	0,41	2
1	16	-0,305	0,017	0,96	0,89	0,47	0,62	1
1	17	0,783	0,017	1,23	1,39	0,30	0,39	2
1	18	-0,344	0,017	0,83	0,74	0,56	0,63	1
1	19	0,459	0,017	0,93	0,90	0,52	0,46	2
1	20	--	--	--	--	--	--	--
1	21	-0,018	0,017	0,84	0,79	0,56	0,56	2
1	22	0,019	0,017	0,90	0,86	0,52	0,55	2
2	1	-2,119	0,026	0,94	0,81	0,36	0,89	1
2	2	-1,434	0,021	0,83	0,64	0,48	0,81	1
2	3	-1,048	0,019	0,81	0,67	0,53	0,76	1
2	4	0,069	0,017	1,09	1,09	0,40	0,54	2
2	5	-0,503	0,017	0,88	0,79	0,51	0,66	1
2	6	-0,483	0,017	0,93	0,96	0,47	0,65	1
2	7	-0,798	0,018	0,97	1,05	0,42	0,71	1
2	8	--	--	--	--	--	--	--
2	9	--	--	--	--	--	--	--
2	10	1,580	0,019	1,23	1,47	0,29	0,25	3
2	11	0,150	0,017	1,01	0,97	0,46	0,52	2
2	12	0,009	0,017	0,96	0,94	0,48	0,55	2
2	13	0,068	0,017	1,03	1,04	0,43	0,54	2
2	14	0,759	0,017	1,02	1,04	0,45	0,40	2
2	15	0,502	0,017	0,93	0,91	0,51	0,45	2
2	16	0,054	0,017	0,95	0,90	0,50	0,54	2
2	17	0,900	0,018	1,16	1,29	0,35	0,37	2
2	18	0,921	0,018	1,17	1,40	0,34	0,37	2
2	19	0,275	0,017	1,03	1,04	0,44	0,50	2
2	20	0,116	0,017	1,03	1,05	0,43	0,53	2
2	21	0,434	0,017	1,01	1,00	0,46	0,47	2
2	22	--	--	--	--	--	--	--

Los ítems número 20 del cuadernillo 1, y 8, 9 y 22 del cuadernillo 2 de Castellano como segunda lengua no muestran un buen ajuste al modelo Rasch, por lo cual no han sido incluidos en la estimación de las medidas de los estudiantes evaluados.

## 5.5 EVIDENCIAS DE LA CONFIABILIDAD DE LAS MEDIDAS

Muñiz (1999) señala que las mediciones en Psicología y Educación deben verse libres de errores de medición. Si las evaluaciones efectuadas con un instrumento son consistentes, carecen de errores de medida y se les considera confiables (Brennan, 2001). La concepción original de confiabilidad se originó en los trabajos de inicios del siglo XX de Charles Spearman (Thomson, 2003) y, en la actualidad, se puede considerar que el objetivo principal de la confiabilidad es tratar de estimar el error existente en las medidas mediante un indicador denominado coeficiente de confiabilidad (Muñiz, 1999).

Muchas veces se habla de la *confiabilidad del test* o se dice «este test es confiable», como si la confiabilidad fuese una propiedad intrínseca del test. La confiabilidad no es una propiedad inmutable del test, sino que es una propiedad que se aplica a las puntuaciones derivadas de aplicar un test, que tendrá un uso específico en un grupo concreto de personas (Sawilowsky, 2003; Thomson, 2003; Thomson y Vacha-Haase, 2003). Es decir, si aplicamos el mismo instrumento en una muestra distinta, encontraremos probablemente un resultado distinto cuando analicemos la confiabilidad de las medidas derivadas de aplicarlo en este grupo de personas.

Wilson (2005) señala que al crear un instrumento se asume que cada persona es medida en la cantidad del constructo o rasgo latente ( $\theta$ ) que se supone debe medir el instrumento. Sin embargo, cuando una persona responde a un instrumento de medida, puede haber muchas otras cosas aparte de  $\theta$  que pueden influir en su respuesta generando un estimado de la habilidad, que es una combinación lineal de la cantidad del rasgo o atributo latente que se desea medir y el error de medición, lo que se representa con la siguiente fórmula:

$$\hat{\theta} = \theta + \varepsilon \quad (5.17)$$

Siguiendo las ideas de Wilson (2005), se pueden identificar cuatro áreas distintas que potencialmente pueden contribuir al error de medición:

- características de la persona: su interés en el tema, estado de ánimo y de salud, etc.,
- condiciones de aplicación: la temperatura y ruido ambiental, el momento del día, etc.,
- características del instrumento: los ítems específicos que lo componen y la forma en que estos son presentados,
- calificación de los ítems: la forma en que los correctores son entrenados la consistencia y precisión de sus calificaciones.

Como se señaló anteriormente, el objetivo de la confiabilidad es tratar de estimar el error existente en las medidas mediante un indicador numérico. Si bien existen diferentes coeficientes de confiabilidad (Brennan, 2001; Haertel, 2006), el tipo de coeficiente de confiabilidad utilizado depende del uso e interpretación propuesta de las puntuaciones o medidas derivadas de la aplicación de un test; por lo tanto, es importante reportar el coeficiente apropiado (Linn, 2006). Al respecto la AERA, APA y NCME (1999) señalan, entre otros, los siguientes estándares:

- Estándar 2.1: para cada puntaje total, subpuntaje o combinación de puntuaciones que se interprete deben reportarse coeficientes de confiabilidad relevantes, errores estándar de medición o funciones de información del test.
- Estándar 2.2: el error estándar de medición, tanto global como condicional (si es relevante), debe ser reportado tanto en las unidades originales del test (puntaje bruto), como en las unidades de cada puntaje derivado que se recomienda utilizar para la interpretación del test.

La confiabilidad de las medidas, derivadas de aplicar las pruebas a los estudiantes que participaron de la ECE, se estimó mediante el índice de confiabilidad de la separación de personas. Esta es una forma de estimación vinculada a la consistencia interna. Ello se debe a que su cálculo se basa en la información referida a la variabilidad contenida en una sola aplicación del instrumento; concretamente, se refiere a la proporción de varianza explicada por el estimador de la medida de habilidad de una persona (Wilson, 2005).

Siguiendo la presentación de Schumacker (2007), Wilson (2005) y Wrigth y Masters, (1982), se expondrá la lógica detrás del cálculo de este coeficiente de confiabilidad:

La varianza total de una medida de habilidad estimada es la siguiente:

$$\text{var}(\hat{\theta}) = \frac{1}{N-1} \sum_{n=1}^N (\hat{\theta} - \bar{\theta})^2 \quad (5.18)$$

Donde  $\bar{\theta}$  es la habilidad promedio estimada para todas las personas ( $N$ ) que responden un test.

La varianza explicada por los errores puede ser calculada como la media cuadrática de los errores estándar de medición (MCE). Esto se debe a que hay un problema si se suman todos los residuos de todas las personas, tal y como lo exponen Bond y Fox (2007). Los residuos en un modelo dicotómico tendrán valores positivos, cuando la respuesta observada es 1, y negativos, cuando es 0. Si se les suma a lo largo de todas las personas, se obtendrá como valor 0. Por ello, se deben elevar los residuos al cuadrado:

$$MCE = \text{var}(\hat{\sigma}) = \frac{1}{N} \sum_{n=1}^N ee(\theta_n)^2 \quad (5.19)$$

Donde  $ee(\theta_n)$  es el error estándar de la estimación de la habilidad de cada persona  $n$  que responde un test.

Por su parte, la varianza explicada por el modelo es la diferencia de las dos varianzas antes señaladas:

$$\text{var}(\theta) = \text{var}(\hat{\theta}) - \text{var}(\hat{\sigma}) \quad (5.20)$$

Finalmente, la proporción de varianza explicada por el modelo, denominada índice de confiabilidad de la separación de las personas, es como sigue:

$$Rp = \frac{\text{var}(\theta)}{\text{var}(\hat{\theta})} \quad (5.21)$$

El denominador de  $Rp$  representa la variabilidad total que hay entre las personas, es decir, cuanto difieren las personas en la medición del constructo de interés. El numerador representa la parte de dicha variabilidad que puede ser reproducida por el modelo Rasch. Muchas veces se le denomina a esta última *variabilidad ajustada de las personas* (Bond y Fox, 2007).

Este coeficiente 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, 2007). El coeficiente sirve para indicarnos la capacidad de las medidas de un test para diferenciar las cantidades de rasgo latente que poseen los evaluados (Wrigth 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, 2007).

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, 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 13** Análisis de confiabilidad de las medidas derivadas de la aplicación de las pruebas

Prueba	$R_p$	$G_p$
Comprensión lectora ECE	0,86	2,45
Comprensión lectora MC	0,86	2,46
Matemática ECE	0,88	2,77
Matemática MC	0,88	2,77
Castellano como segunda lengua	0,88	2,74

Considerando el valor de  $R_p$  de las distintas pruebas aplicadas en la ECE, podemos apreciar que, en el peor de los casos, la varianza de error es del 16%, por lo cual se puede afirmar que las medidas derivadas de aplicar dichas pruebas poseen adecuadas evidencias de confiabilidad.

Otra manera de aproximarnos al error de medición es mediante el error estándar de las medidas de habilidad estimadas  $ee(\theta)$ , que puede ser utilizado para construir intervalos de confianza de  $\theta$  (Wright y Stone, 1982). A diferencia de la propuesta de la Teoría Clásica de los Test (Charter y Feldt, 2002),  $ee(\theta)$  varía dependiendo de la localización de la persona en el continuo de  $\theta$  y la relación entre ambos típicamente tiene la forma de una U (Wilson, 2005). Otra manera de expresar esta relación es mediante la *Información* que es igual al recíproco del cuadrado de  $ee(\theta)$ :

$$Inf(\theta) = \frac{1}{[ee(\theta)]^2} \quad (5.22)$$

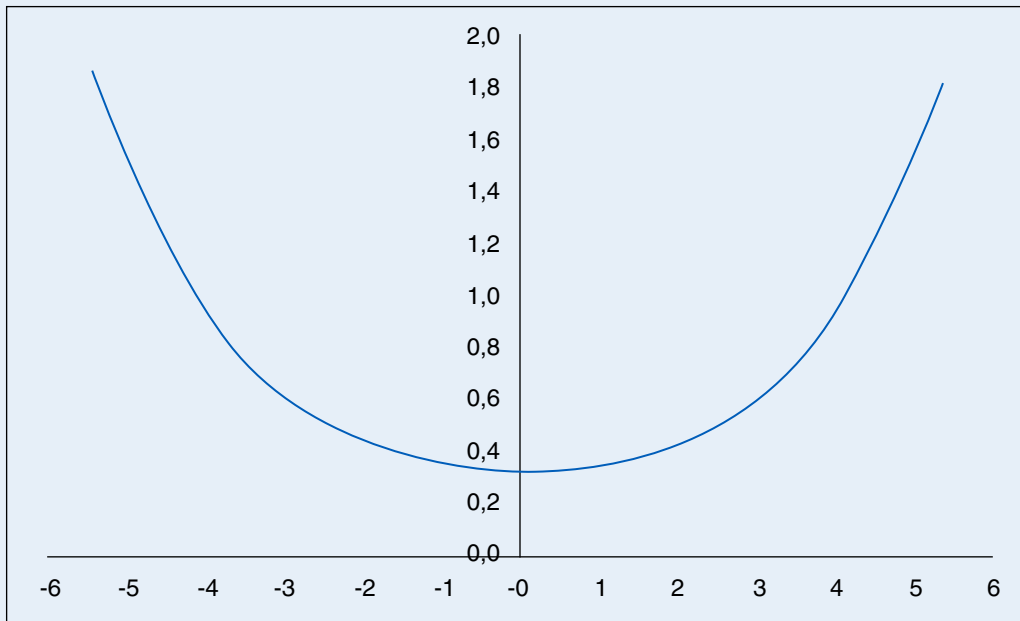
Es importante señalar que la información de un test es igual a la suma de la información de cada ítem:

$$Inf(\theta) = \sum_{i=1}^I Inf_i(\theta) \quad (5.23)$$

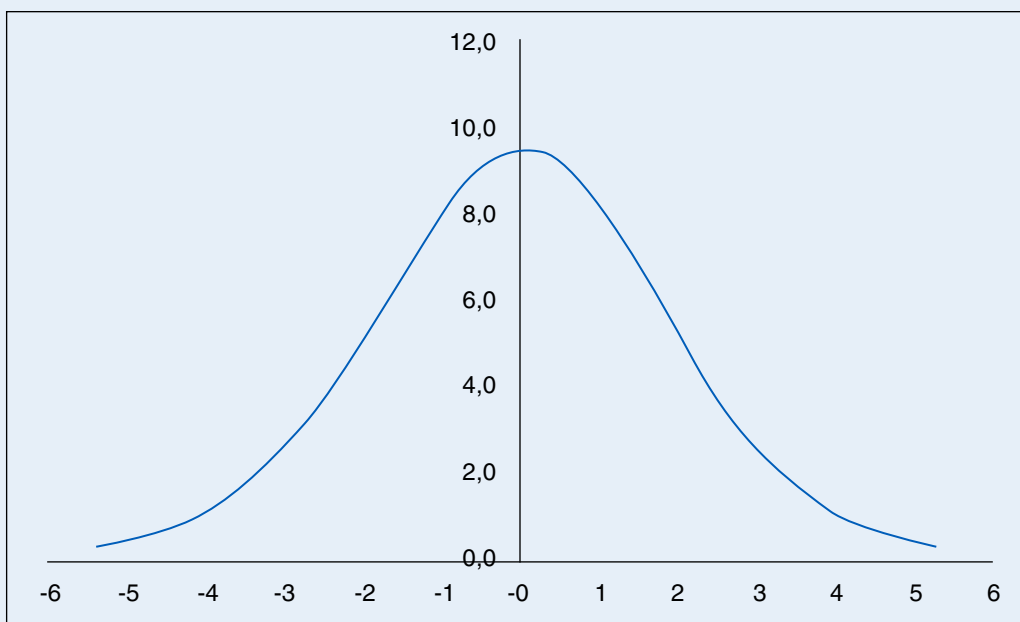
Una característica importante del error estándar de medición es cuando una muestra y un test no son concordantes (el test en general es muy fácil o muy difícil para esa muestra) con los errores estándar para esa muestra, son más grandes que en una muestra cuya habilidad es concordante con la dificultad del test (Wright, 1991).



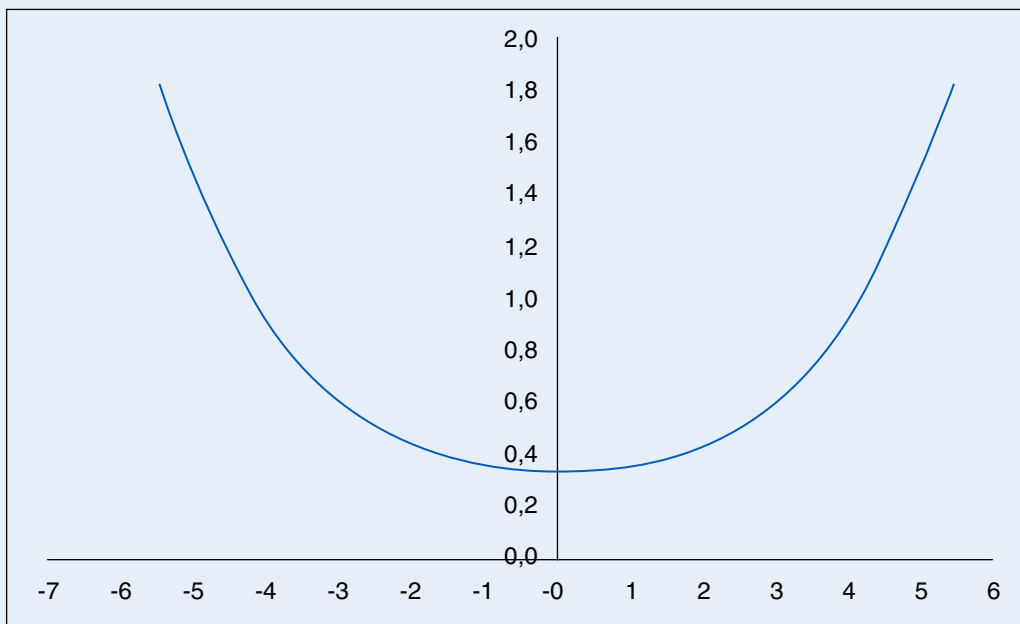
Se presentan, a continuación, las figuras del error estándar y la función de información para todas las pruebas aplicadas, centrada en 0:



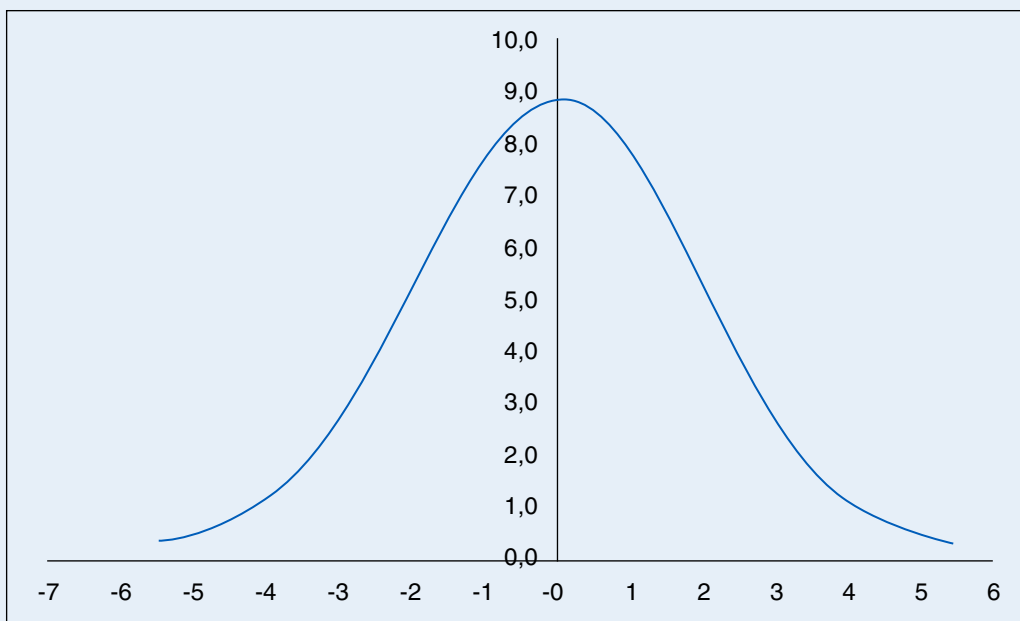
**Figura 6.** Relación entre el error estándar de medición y la habilidad estimada, Comprensión lectora ECE



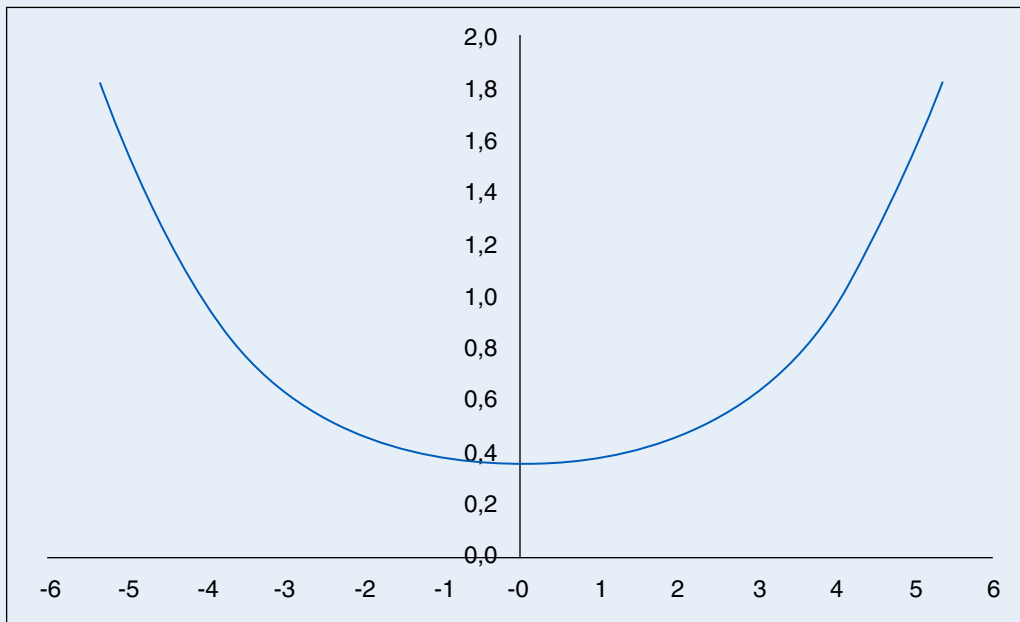
**Figura 7.** Función de información, Comprensión lectora ECE



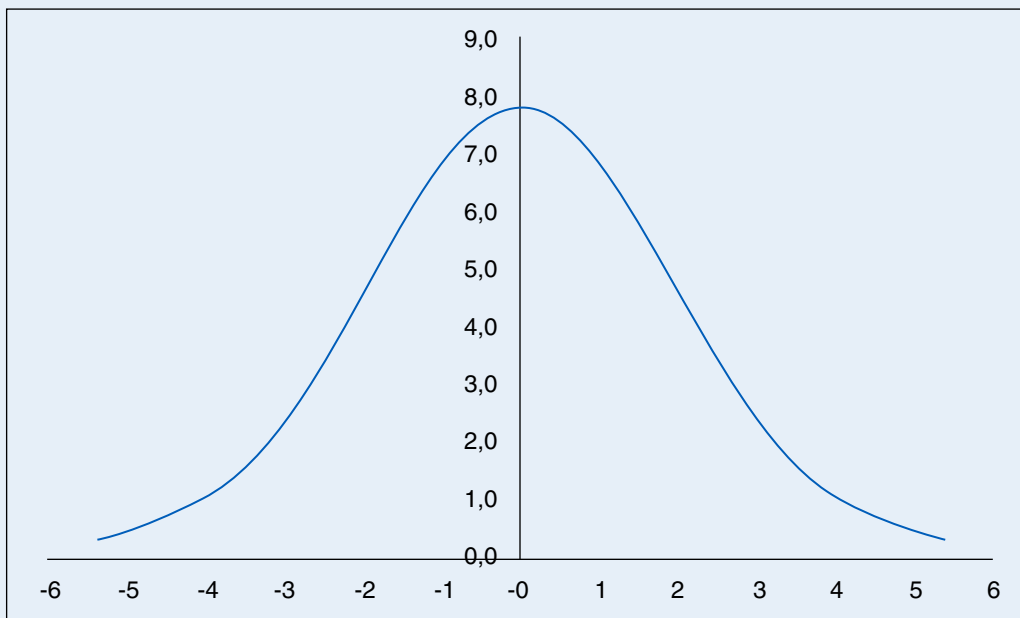
**Figura 8.** Relación entre el error estándar de medición y la habilidad estimada, Comprensión lectora MC



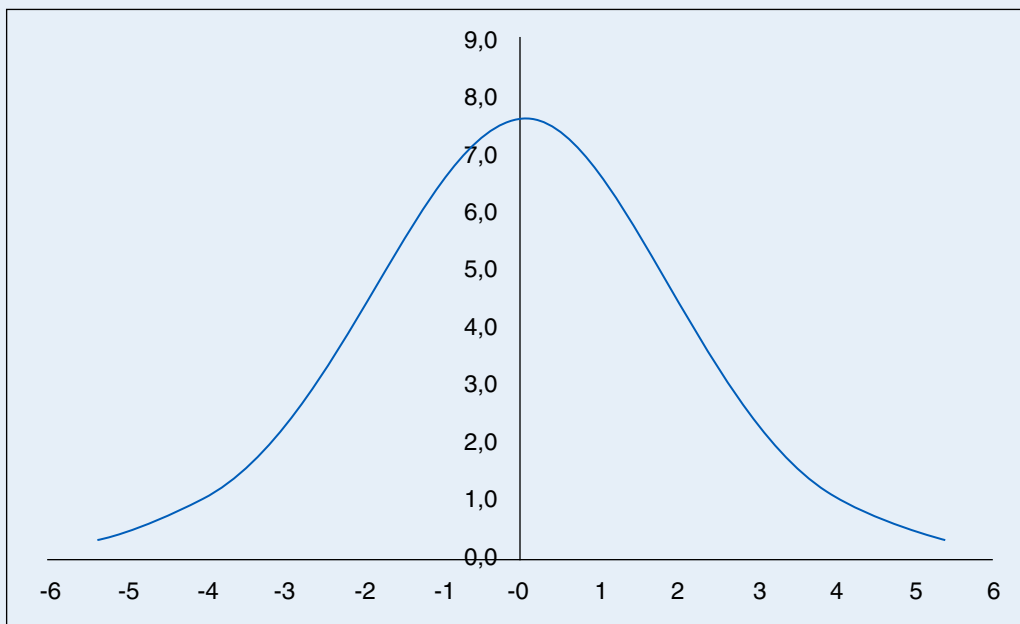
**Figura 9.** Función de información, Comprensión lectora MC



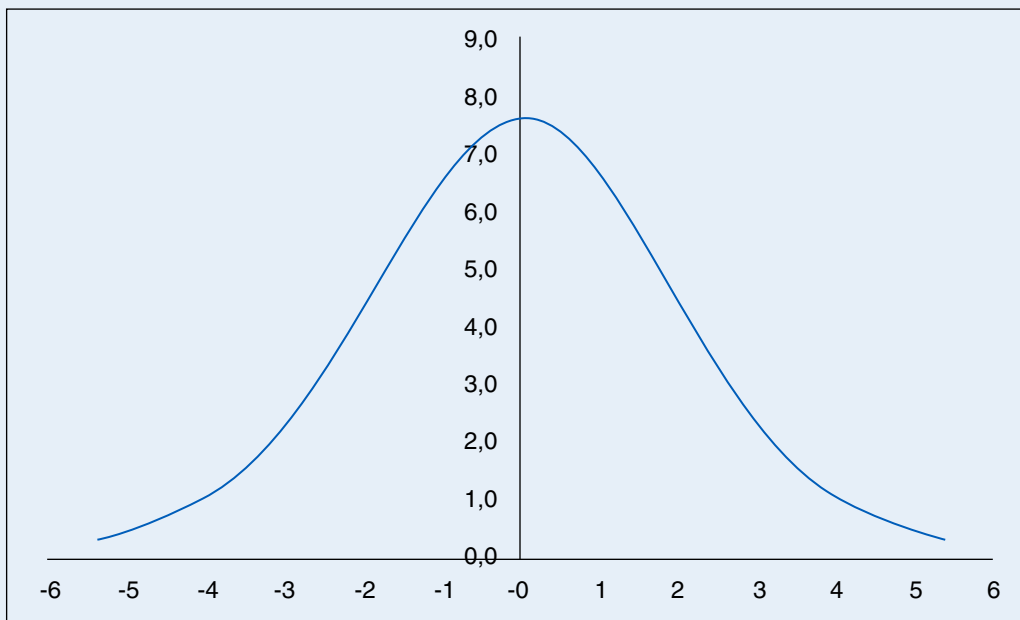
**Figura 10.** Relación entre el error estándar de medición y la habilidad estimada, Matemática ECE



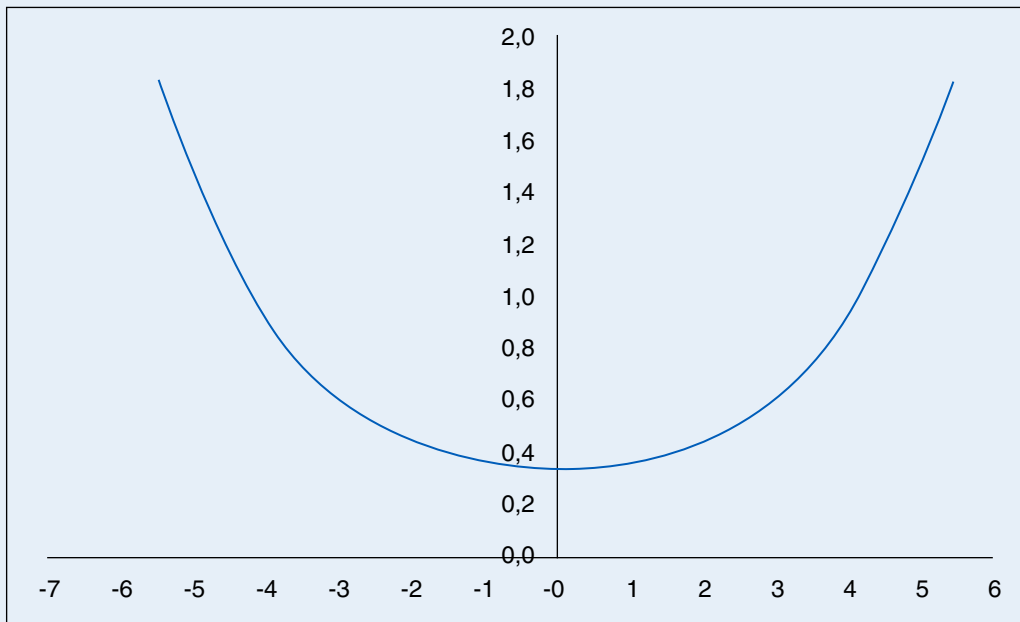
**Figura 11.** Función de información, Matemática ECE



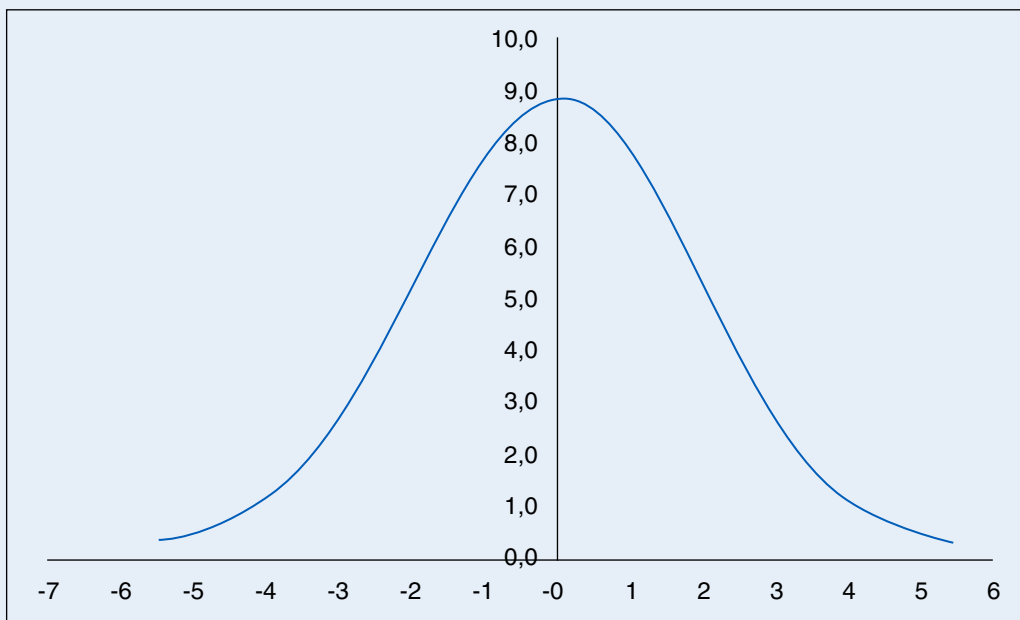
**Figura 12.** Relación entre el error estándar de medición y la habilidad estimada, Matemática MC



**Figura 13.** Función de información, Matemática MC



**Figura 14.** Relación entre el error estándar de medición y la habilidad estimada, Castellano como segunda lengua



**Figura 15.** Función de información, Castellano como segunda lengua

## 5.6 EVIDENCIAS DE VALIDEZ

El concepto de validez no ha sido entendido siempre de la misma manera a lo largo de la historia de la psicometría. Kane (2001) propone tres etapas en su desarrollo histórico: el modelo basado en el criterio, el modelo del constructo y la aproximación moderna. Tal vez, una de las perspectivas más conocidas sobre la validez sea la de Cronbach y Meehl (1955), quienes señalan que esta implica acumular evidencia en tres categorías básicas: el criterio, el contenido y el constructo. Si bien los autores antes mencionados señalan que la *validez de constructo* incluye dentro de sí a las otras dos, muchos autores identifican tres tipos de evidencias de validez (Muñiz, 1999).

La aproximación moderna se debe a la propuesta de Messick de 1989 (Kane, 2006). Este autor considera que la aproximación tradicional es fragmentada e incompleta, y critica duramente la visión tripartita de la validez, a la que considera como un concepto unitario. Presenta la validez como un juicio evaluativo integral del grado en el cual la evidencia empírica y teórica soportan lo adecuado y apropiado de las interpretaciones y acciones basadas en los puntajes de una prueba u otra forma de evaluación.

Siguiendo esta aproximación moderna, la AERA, APA y NCME (1999) señalan en sus estándares que la validez se refiere al grado en el cual teoría y evidencias sustentan las interpretaciones de las puntuaciones implicadas en los usos propuestos del test. Además, en esos mismos estándares, se diferencian cinco fuentes de evidencias de validez:

- 1) Contenido: congruencia entre el contenido del test y el dominio que se supone debe medir.
- 2) Proceso de respuesta: consistencia entre las actividades que el test demanda a los examinados y el proceso de respuesta que se supone representa.
- 3) Estructura interna: consistencia entre la estructura del constructo y las relaciones entre los ítems o subescalas del test.
- 4) Relaciones con otras variables:
  - a) Evidencias convergentes y discriminantes: relaciones entre el test y otras medidas de constructos similares o diferentes.
  - b) Relaciones test-criterio: relación existente entre el test y otra medida del constructo (estudios concurrentes); o entre el test y otros constructos que pueden ser predichos a partir de las puntuaciones o medidas del test (estudios predictivos).
  - c) Generalización de la validez: estudia cómo las correlaciones entre las puntuaciones de un test y los criterios pueden diferir entre grupos.
- 5) Consecuencias de los test: evidencias referidas a los efectos de la evaluación. Implica el análisis de los efectos positivos y negativos vinculados a las decisiones basadas en el uso de los test.

El proceso mediante el cual se recolectan evidencias sobre la interpretación y usos de los test se denomina *validación* e implica, además, valorar la factibilidad de dichas interpretaciones y usos propuestos (Kane, 2006). Las evidencias de validez, referidas a las inferencias que se realizan sobre la base de las medidas derivadas de aplicar las pruebas ECE, se circunscriben a dos de los diferentes tipos de evidencias: las vinculadas al contenido de las pruebas y las referidas a su estructura interna.

### 5.6.1 EVIDENCIAS DE VALIDEZ REFERIDAS AL CONTENIDO DE LAS PRUEBAS

Este tipo de evidencias de validez se vinculan a la necesidad de garantizar que el test constituye una muestra adecuada y representativa de los contenidos que se pretende evaluar con él (McGartland, Berg-Weger, Tebb, Lee y Rauch, 2003; Muñiz, 1999). En esta misma línea, Suen (1990) indica que los ítems en un test, supuestamente, constituyen una muestra representativa de todos los ítems que se pueden emplear para medir el constructo de interés. Si este supuesto se cumple, entonces, se podrá decir que la prueba tiene adecuadas evidencias de validez referidas a su contenido, quedando esta definida como la representatividad de los ítems empleados en un test para medir un constructo particular. La representatividad de un test puede definirse como la

precisión con que podemos hacer inferencias acerca de la puntuación de la persona en el universo de indicadores, a partir de la puntuación en la muestra de ítems que constituyen la prueba (Paz, 1996). Esto es posible gracias a que la muestra de ítems en la prueba reproduce las características esenciales del universo de ítems relevantes, destinados a medir cierto constructo en la proporción y en el balance adecuados (Suen, 1990).

Muñiz (1999) señala que la práctica más usual para lograr este tipo de validez consiste en enumerar todas las áreas de contenido que se consideren importantes o imprescindibles y, luego, asegurarse que la prueba contenga ítems que hagan referencia a cada una de ellas en una proporción adecuada. En ese sentido, el proceso de validación implica al mismo proceso de construcción, pues se debe seguir una tabla de especificaciones. Las pruebas de la ECE se construyen sobre la base del *Marco de trabajo* (UMC, 2009) en el que se encuentra la tabla de especificaciones elaborada por la Unidad de Medición de la Calidad Educativa. En el *Marco de trabajo*, se explica el enfoque de cada área; los objetivos y objetos (variables) de la evaluación; las características de la población evaluada; las competencias, las capacidades e indicadores evaluados; el modelo y los contextos de evaluación; y, en el caso específico de Comunicación, los tipos y géneros textuales utilizados en la evaluación. En las tablas de especificaciones, se precisa la cantidad y la proporción asignada a cada capacidad, indicador y descriptor a evaluar. En el caso de Comunicación, también se precisa la cantidad y proporción de los diferentes tipos y géneros textuales. Es importante señalar que tanto el *Marco de trabajo* como las tablas de especificaciones guardan correspondencia con la última versión de los *Mapas de Progreso* propuestos por el Instituto Peruano de Evaluación, Acreditación y Certificación de la Calidad de la Educación Básica (Ipeba, 2013) y con los lineamientos del *Diseño Curricular Nacional* (Minedu, 2008) y con las *Rutas del Aprendizaje* (Minedu, 2013) publicados por el Ministerio de Educación.

En general, para Suen (1990), el proceso de validación de contenido implica la definición del universo de indicadores y sus respectivos ítems representativos; la identificación de jueces competentes dentro del área que el test pretende medir; y el juicio por parte de estos expertos respecto de la relevancia y la representatividad de cada ítem empleado para medir el constructo o los constructos implicados en el test que se está construyendo. Este procedimiento es denominado criterio de jueces y es aplicado de manera práctica a diferentes instrumentos, como lo ejemplifica, entre otros, el trabajo de George, Schkade e Ishee (2004) y el de Head et. al. (2004). Ecurra (1988) indica que la modalidad más común para obtener la validez de contenido mediante el criterio de jueces es solicitar la aprobación o desaprobación de un ítem de la prueba por parte de varios jueces, cuyo número varía según los propósitos del autor del instrumento. Según lo establecido por la UMC, el proceso de validación incluye que un conjunto de expertos en las áreas evaluadas emita un juicio respecto de la construcción, la correspondencia con el indicador y la dificultad de la tarea de cada ítem empleado para medir el constructo de cada prueba aplicada.

Después de que los equipos de Matemática, Comunicación y de Educación Intercultural Bilingüe han construido los ítems, estos son revisados por expertos de la UMC. Ellos evalúan aspectos como la calidad, actualidad 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 las alternativas, tanto en sus aspectos formales como en su eficacia para la medición del constructo. También se toman en cuenta posibles sesgos socioeconómicos, culturales y de género en la construcción. En el caso específico de Comunicación (segundo grado y cuarto grado EIB), se evalúa también la pertenencia de los textos a los tipos y géneros textuales que se señalan, la densidad de información, el vocabulario, la sintaxis, la familiaridad del tema, la extensión y el formato. Solo los ítems que, según los revisores, cumplen con estos criterios pasan a una aplicación piloto luego de la cual se analiza su ajuste al modelo Rasch.

Luego de seleccionar los ítems mediante la piloto y el análisis posterior, se organizan en cuadernillos para que sean validados por expertos de la Dirección de Educación Primaria (DEP), de la Dirección General de Educación Intercultural Bilingüe y Rural (Digeibir) y del Instituto Peruano de Evaluación, Acreditación y Certificación de la Calidad de la Educación Básica (Ipeba). En esta segunda validación de expertos, se solicita a los jueces que evalúe los aspectos mencionados mediante una ficha técnica (véase anexos 6, 7 y 8). Esta ficha tiene cuatro secciones: tres en que se presentan indicadores que evalúan las características generales del ítem, así como las características específicas del enunciado y de las alternativas, y una para que el juez escriba sugerencias puntuales sobre el ítem. Las fichas presentan ciertas particularidades, según el ítem esté incluido en las pruebas que evalúan los constructos de comprensión lectora o de la comprensión del número, de las operaciones y del Sistema de Numeración Decimal. En el caso específico de Comunicación, también se incluyen indicadores referidos al texto. Cada experto completa la ficha y formula las observaciones que, según 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, en Comunicación, al texto.

### 5.6.2 EVIDENCIAS DE VALIDEZ REFERIDAS 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 interna, dado el constructo de interés. En este sentido, un tipo de evidencias tiene que ver 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 dadas a él son producidas sobre la base de un único atributo. Wright y Linacre (1998) 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, Swaminathan y Rogers, 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).

Al aplicar el análisis factorial se encuentra un problema importante en la factorización de ítems dicotómicos. Waller (2001) nos dice que el análisis factorial asume un modelo lineal en la regresión de la respuesta a un ítem sobre el puntaje factorial subyacente. El problema está en que, al momento de aplicar el modelo lineal con datos binarios, las personas con puntajes factoriales muy altos o muy bajos pueden tener probabilidades superiores a 1.00 o inferiores a 0.00, de «acertar» o «estar de acuerdo» con el ítem. Por otro lado, Cuesta (1996) nos dice que en el caso de que la variable sea dicotómica al momento de realizar un análisis factorial sobre una matriz de correlaciones, se aplica una modificación del coeficiente de Pearson, conocida como coeficiente *phi*. En estos casos, generalmente, se forman factores espurios. Nunnally y Berstein (1995) señalan que los ítems fáciles se juntarán con otros fáciles y los difíciles, con otros difíciles; además, los de alto endosamiento formarán factores separados de los de bajo endosamiento, a pesar que los ítems en el fondo sean unidimensionales. Por este motivo, se les suele llamar factores de dificultad (Cuesta, 1996; Kubinger, 2003).

En el caso de las pruebas aplicadas en la ECE, se realizó un análisis de componentes principales de los residuos estandarizados, luego de ajustar el modelo Rasch unidimensional (Linacre, 2005). 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, 2007); 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 sirven 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 14** Análisis de unidimensionalidad de las medidas derivadas de la aplicación de las pruebas

Prueba	Varianza de la dimensión principal (%)	Varianza modelada (%)	Primer autovvalor	Porcentaje del primer autovvalor
Comprensión lectora ECE	27,3	27,8	2,2	3,6
Comprensión lectora MC	29,6	30,2	2,3	3,7
Matemática ECE	35,9	35,5	2,0	3,3
Matemática MC	36,7	36,3	1,9	3,2
Castellano como segunda lengua	28,3	28,4	2,6	4,7



Linacre (2005) señala 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 3,0, 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 serias evidencias en contra del supuesto de unidimensionalidad del conjunto de ítems incluidos en estos instrumentos de medición.

## 5.7 EQUIPARACIÓN DE PUNTUACIONES

En el marco del análisis Rasch, las puntuaciones del rasgo latente tienen un origen y escala de medición arbitraria (Bond y Fox, 2007). Por ello, dos instrumentos que miden lo mismo, aplicados a distintas muestras de personas, no se encontrarán en la misma métrica. A fin de poder establecer las comparaciones necesarias, se debe 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 (Livingsstone, 2004).

Diversas síntesis teóricas que tratan sobre la equiparación de puntuaciones (Kolen y Brennan, 2004; Navas, 1996) citan los trabajos de Angoff y los de 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 última es tener una métrica común para dos o más medidas de un mismo rasgo, de manera que se puedan comparar los resultados de personas a las que se les aplican diferentes instrumentos que miden el mismo rasgo.

Por ejemplo, supongamos que tenemos dos test X e Y. Queremos equiparar al 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. De este modo buscaremos resolver:

$$Y^* = f(X) \quad (5.24)$$

Fórmula en la que  $Y^*$  son las puntuaciones de X equiparadas a la métrica de Y.

El desarrollo del sistema de equiparación implica el uso de diseños que permiten luego obtener  $Y^* = f(X)$ . Revisando la literatura sobre el tema (Ho y Osborn, 2005; Kolen y Brennan, 2004; Livingsstone, 2004; Navas, 1996; Zhu, 1998), podemos 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 el utilizado para colocar las medidas de las pruebas de las distintas evaluaciones censales de segundo grado de primaria en la misma métrica que las pruebas aplicadas en el año 2007 por la Unidad de Medición de la Calidad Educativa del Ministerio de Educación del Perú.

En este diseño, se trabaja con dos muestras de personas, que no necesariamente han sido extraídas de la misma población. En cada grupo, 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 a equiparar. Es por ello que muchas veces se les denomina ítems de anclaje o *test de anclaje*. Además, es importante considerar que el *anclaje* puede ser de dos tipos:

- Anclaje interno: los ítems comunes aparecen intercalados o formando un bloque junto con el resto de ítems de la prueba y se les considera en el cálculo del puntaje total. Generalmente, a esta variante se le denomina ítems de anclaje.
- Anclaje externo: los ítems comunes aparecen formando un test independiente y no se les considera en el cálculo del puntaje total. A esta variante suele denominarse *test de anclaje* y no es utilizada en el marco de la Teoría de Respuesta al Ítem, mas sí en el de la Teoría Clásica de los Test.

Una vez obtenidos los datos, mediante algún diseño de equiparación, es necesario aplicar distintos métodos

para colocar ambas puntuaciones o medidas en la misma métrica. Para ello, existen diversos procedimientos. En esta ocasión, se utilizó la transformación lineal: se consideran equivalentes las puntuaciones directas que corresponden a la misma puntuación típica. En el contexto de la TRI, se denomina a este método *mean / sigma* (véase Kolen y Brennan, 2004), e implica el cálculo de dos constantes de equiparación *a* y *b*, de tal manera que surge la siguiente fórmula:

$$medida_{equiparadaECE} = b + a \times medida_{original} \quad (5.25)$$

De la que resultan:

$$a = \frac{S_{medidaECE}}{S_{medida\_original}} \quad (5.26)$$

$$b = \bar{X}_{medidaECE} - a \times \bar{X}_{medida\_original} \quad (5.27)$$

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 de mantenerse constante<sup>6</sup>, independientemente del grupo poblacional en el que se le aplique. En caso de no cumplirse con este supuesto, se dice que el ítem presenta un funcionamiento diferencial (FDI) a lo largo del tiempo.

Una posibilidad para analizar el FDI es la señalada por Bond y Fox (2001), que consiste en el uso del método gráfico basado en el diagrama de dispersión y los errores estándar de la dificultad de los ítems. Este método radica en obtener de forma independiente las dificultades de los ítems comunes en cada una de las muestras en las que fueron aplicados. Luego se llevan estos ítems a un plano de coordenadas que muestra en el eje X las dificultades obtenidas en la población uno. En el eje Y aparecen las dificultades calculadas para la población dos. Al forzar una regresión con una pendiente de uno, se puede ver qué tanto se alejan los ítems de esa línea de regresión y establecer cuáles son los que mejor se ajustan al supuesto de invarianza. Si los ítems se alejan mucho de la línea de ajuste, están funcionando de forma distinta en ambas poblaciones (Hambleton, Swaminathan y Rogers, 1991; Ingebo, 1997).

Una pregunta que nos podemos plantear es qué tanto se puede alejar un ítem de esta línea de ajuste antes de romper con el supuesto de invarianza. Para ello, se pueden trazar unas líneas de control basadas en el error estándar del ítem<sup>7</sup> (Bond y Fox, 2001; Luppescu, 1991). Es importante señalar que las líneas de control no son rectas, sino curvas. Esto se debe a que el error estándar es menor hacia el centro de la distribución de dificultades y mayor en los extremos inferior y superior. Si un ítem cae fuera de dichas líneas de control, diremos que no cumple con los requerimientos de invarianza y, por lo tanto, presenta funcionamiento diferencial entre ambas poblaciones. Este método fue el empleado para analizar el FDI de los ítems de la ECE.

Como se mencionó al inicio de este reporte técnico, uno de los objetivos de la Evaluación Censal de Estudiantes 2013 está vinculado con la necesidad de establecer comparaciones entre los resultados obtenidos en la ECE 2012 con el propósito de medir cambios en los logros de aprendizajes de los estudiantes. Estas comparaciones son realizadas mediante una muestra de control (MC).

La comparación entre los resultados de la ECE 2013 y la ECE 2012 es posible gracias a que se realizó un proceso de equiparación de las pruebas aplicadas en los distintos años. La manera como se coloca en la misma escala las pruebas aplicadas en el marco de las diferentes evaluaciones censales implicó los siguientes pasos:

- A una muestra representativa de estudiantes a escala nacional se les aplicó en el año X una prueba conformada por los ítems que serán utilizados en la evaluación censal del año X+1. Utilizando el método de equiparación por personas en común, se colocaron ambas pruebas en la misma métrica

<sup>6</sup> En realidad, la dificultad del ítem debe mantenerse dentro de ciertos márgenes de error al comparar su dificultad en dos poblaciones diferentes.

<sup>7</sup> Cuando se estima la dificultad de un ítem, en realidad, solo nos aproximamos a su valor. No lo podemos conocer exactamente. Siempre tiene asociado cierto error de estimación (error estándar). En función a este error estándar, se pueden trazar intervalos dentro de los cuales es probable que se encuentre la verdadera dificultad del ítem.

en la que el punto de referencia fue la evaluación del año X. Este procedimiento es denominado *primera etapa* en este documento.

- Se utilizaron en el año X+1 los parámetros de dificultad calibrados en el año anterior con la finalidad de identificar aquellos ítems que no mostraban un marcado funcionamiento diferencial para usarlos como ítems en común. Es decir, se empleó el diseño de equiparación por ítems comunes a fin de calcular las constantes de equiparación lineal utilizadas para colocar los ítems del año X+1 en la métrica del año X. A este proceso le hemos denominado *segunda etapa*.
- Este mismo proceso se replicará, es decir, se aplicará para las siguientes evaluaciones: X+1 y X+2 con la finalidad de ponerlas en la misma métrica, al igual que con X+2 y X+3, siendo siempre el punto de referencia el año X.

Finalmente, se presentan las tablas con los resultados de la equiparación por ítems comunes de las cinco pruebas utilizadas en la ECE 2013. En la tabla 11, se muestra la cantidad de ítems en común en las pruebas aplicadas. Las pruebas MC y castellano como segunda lengua contienen ítems de años anteriores. Además, las pruebas ECE tienen ítems en común con las pruebas MC, es decir, las pruebas MC se equiparan con el año anterior y las pruebas ECE son equiparadas con las pruebas MC que ya se encuentran en la métrica del año anterior.

**Tabla 15** Número de ítems programados y utilizados para la equiparación de las medidas

Prueba	Programados	Usados
Comprensión lectora ECE	39	33
Comprensión lectora MC	27	19
Matemática ECE	34	26
Matemática MC	28	16
Castellano como segunda lengua	37	22

**Tabla 16** Análisis DIF de las medidas de Comprensión lectora ECE usadas en la equiparación 2013 con 2012

Ítem	2012		2013		Trans.	DIF	
	Medida	Error	Medida	Error	medida	Medida	Error
cl1_01	-1,199	0,075	-1,651	0,071	-1,240	0,052	0,103
cl1_02	-0,550	0,061	-1,192	0,062	-0,731	0,191	0,087
cl1_03	-0,510	0,061	-0,929	0,058	-0,440	-0,062	0,084
cl1_04	-1,123	0,073	-1,708	0,073	-1,303	0,191	0,103
cl1_05	-0,189	0,056	-0,785	0,056	-0,280	0,099	0,079
cl1_07	-0,476	0,060	-0,889	0,057	-0,038	-0,072	0,083
cl1_08	0,439	0,049	-0,157	0,049	-0,396	0,029	0,069
cl1_09	0,231	0,051	-0,366	0,051	0,415	0,053	0,072
cl1_10	1,838	0,043	1,137	0,043	0,184	-0,010	0,061
cl1_11	0,798	0,046	0,240	0,046	1,849	-0,053	0,065
cl1_12	0,353	0,050	-0,124	0,048	0,855	-0,094	0,069
cl1_13	0,983	0,045	0,438	0,045	0,452	-0,088	0,064
cl1_14	0,908	0,046	0,283	0,045	1,074	0,009	0,064
cl1_15	1,527	0,043	0,911	0,043	0,903	-0,070	0,061
cl1_16	1,290	0,044	0,511	0,044	1,598	0,138	0,062
cl1_17	1,000	0,045	0,277	0,046	1,155	0,108	0,064
cl1_18	1,183	0,044	0,462	0,045	0,896	0,085	0,063
cl1_19	1,718	0,043	0,951	0,043	1,101	0,076	0,061
cl1_20	0,168	0,052	-0,457	0,052	1,642	0,092	0,074
cl1_21	2,347	0,043	1,525	0,043	0,083	0,067	0,061
cl1_23	2,549	0,044	1,896	0,044	2,278	-0,143	0,062
cl2_02	-0,381	0,054	-0,775	0,055	2,689	-0,103	0,077
cl2_03	-0,998	0,061	-1,339	0,065	-0,269	-0,094	0,089
cl2_04	-1,387	0,067	-1,688	0,072	-0,894	-0,095	0,098
cl2_06	-0,663	0,057	-1,037	0,059	-1,281	-0,095	0,082
cl2_07	0,230	0,049	-0,280	0,050	-0,560	-0,043	0,070
cl2_13	1,072	0,046	0,437	0,045	0,279	0,002	0,064
cl2_14	0,594	0,047	0,057	0,047	1,073	-0,054	0,066
cl2_15	0,009	0,051	-0,363	0,051	0,652	-0,172	0,072
cl2_16	1,043	0,046	0,447	0,045	0,187	-0,039	0,064
cl2_17	1,439	0,045	0,667	0,044	1,084	0,114	0,063
cl2_18	0,564	0,047	0,037	0,047	1,328	-0,062	0,066
cl2_20	0,806	0,046	0,161	0,046	0,630	0,043	0,065

**Tabla 17** Análisis DIF de las medidas de Comprensión lectora MC usadas en la equiparación 2013 con 2012

Ítem	2012		2013		Trans. medida	DIF	
	Medida	Error	Medida	Error		Medida	Error
cl1_01	-1,579	0,069	-1,784	0,075	-1,272	-0,307	0,102
cl1_02	-0,726	0,057	-1,136	0,061	-0,585	-0,142	0,083
cl1_07	-0,514	0,054	-1,061	0,060	-0,505	-0,009	0,081
cl1_14	0,841	0,046	0,322	0,046	0,963	-0,122	0,065
cl1_15	1,706	0,044	0,942	0,043	1,621	0,085	0,062
cl1_17	1,109	0,045	0,415	0,045	1,062	0,048	0,064
cl2_01	-1,356	0,009	-2,056	0,084	-1,561	0,205	0,084
cl2_02	-1,053	0,008	-1,556	0,071	-1,030	-0,023	0,071
cl2_04	-0,699	0,007	-1,408	0,068	-0,873	0,174	0,068
cl2_05	-0,453	0,007	-0,952	0,059	-0,389	-0,064	0,059
cl2_07	-0,883	0,008	-1,464	0,069	-0,933	0,050	0,069
cl2_08	-0,596	0,007	-1,193	0,063	-0,645	0,049	0,063
cl2_09	-0,258	0,006	-0,934	0,059	-0,370	0,112	0,059
cl2_10	1,048	0,005	0,395	0,046	1,040	0,008	0,046
cl2_11	0,864	0,005	0,258	0,047	0,895	-0,031	0,047
cl2_12	0,693	0,005	0,086	0,048	0,712	-0,019	0,048
cl2_15	0,715	0,005	0,214	0,047	0,848	-0,133	0,047
cl2_20	1,045	0,046	0,209	0,047	0,843	0,202	0,066
cl2_21	2,010	0,045	1,388	0,043	2,094	-0,085	0,062
cl1_21	2,347	0,043	1,525	0,043	0,083	0,067	0,061

**Tabla 18** Análisis DIF de las medidas de Matemática ECE usadas en la equiparación 2013 con 2012

Ítem	2012		2013		Trans. medida	DIF	
	Medida	Error	Medida	Error		Medida	Error
ma1_01	-2,111	0,075	-2,118	0,071	-2,094	-0,017	0,103
ma1_02	-1,234	0,057	-1,215	0,058	-1,222	-0,012	0,081
ma1_04	-0,432	0,048	-0,389	0,048	-0,425	-0,007	0,068
ma1_05	-0,774	0,051	-0,741	0,052	-0,765	-0,009	0,073
ma1_06	-0,780	0,051	-0,747	0,050	-0,771	-0,009	0,071
ma1_07	0,438	0,045	0,508	0,045	0,440	-0,002	0,064
ma1_08	0,202	0,046	0,265	0,046	0,206	-0,004	0,065
ma1_09	-0,475	0,048	-0,434	0,048	-0,468	-0,007	0,068
ma1_11	-0,118	0,047	-0,065	0,047	-0,113	-0,005	0,066
ma1_12	1,395	0,047	1,494	0,046	1,392	0,003	0,066
ma1_13	0,384	0,045	0,452	0,045	0,386	-0,002	0,064
ma1_14	-0,763	0,051	-0,729	0,051	-0,754	-0,009	0,072
ma1_15	-0,421	0,048	-0,378	0,048	-0,415	-0,006	0,068
ma1_16	2,189	0,053	2,312	0,051	2,181	0,008	0,074
ma1_17	0,289	0,045	0,354	0,045	0,292	-0,003	0,064
ma1_18	0,747	0,045	0,826	0,045	0,748	-0,001	0,064
ma1_19	0,661	0,045	0,737	0,045	0,661	0,000	0,064
ma1_20	0,046	0,046	0,104	0,046	0,051	-0,005	0,065
ma1_21	0,816	0,045	0,897	0,045	0,816	0,000	0,064
ma2_04	-0,713	0,050	-0,816	0,050	-0,837	0,125	0,071
ma2_05	-1,433	0,056	-1,488	0,058	-1,486	0,053	0,081
ma2_06	-1,854	0,063	-1,714	0,061	-1,704	-0,150	0,088
ma2_10	-0,720	0,050	-0,676	0,049	-0,702	-0,018	0,070
ma2_11	-1,374	0,056	-1,491	0,058	-1,489	0,115	0,081
ma2_18	0,549	0,046	0,551	0,045	0,482	0,067	0,064
ma2_21	0,667	0,046	0,852	0,045	0,772	-0,106	0,064

**Tabla 19** Análisis DIF de las medidas de Matemática MC usadas en la equiparación 2013 con 2012

Ítem	2012		2013		Trans. medida	DIF	
	Medida	Error	Medida	Error		Medida	Error
ma1_04	-0,510	0,048	-0,568	0,048	-0,385	-0,125	0,068
ma1_05	-0,678	0,049	-0,967	0,051	-0,737	0,059	0,071
ma1_06	-0,676	0,049	-0,974	0,051	-0,743	0,068	0,071
ma1_07	0,729	0,046	0,447	0,045	0,512	0,216	0,064
ma1_14	-0,563	0,048	-0,954	0,051	-0,726	0,163	0,070
ma1_15	-0,178	0,047	-0,556	0,048	-0,374	0,195	0,067
ma1_17	0,213	0,046	0,273	0,045	0,359	-0,145	0,064
ma1_18	0,683	0,046	0,808	0,045	0,832	-0,149	0,064
ma2_05	-1,570	0,060	-1,785	0,060	-1,460	-0,110	0,085
ma2_06	-1,339	0,059	-1,598	0,059	-1,295	-0,044	0,083
ma2_07	0,049	0,047	0,049	0,047	0,161	-0,112	0,066
ma2_08	-0,321	0,049	-0,559	0,049	-0,377	0,056	0,069
ma2_12	0,593	0,046	0,446	0,046	0,512	0,081	0,065
ma2_16	-0,740	0,050	-0,844	0,050	-0,628	-0,112	0,071
ma2_18	0,135	0,047	-0,103	0,047	0,026	0,109	0,066
ma2_19	0,662	0,046	0,786	0,046	0,812	-0,150	0,065

**Tabla 20** Análisis DIF de las medidas de Castellano como segunda lengua usadas en la equiparación 2013 con 2012

Ítem	2012		2013		Trans. medida	DIF	
	Medida	Error	Medida	Error		Medida	Error
L2_1_03	-1,725	0,065	-1,777	0,022	-1,826	0,101	0,069
L2_1_05	-0,734	0,054	-0,734	0,018	-0,832	0,097	0,057
L2_1_06	-1,386	0,060	-1,346	0,020	-1,415	0,029	0,063
L2_1_07	0,045	0,052	0,177	0,017	0,037	0,008	0,055
L2_1_08	0,282	0,052	0,485	0,017	0,330	-0,048	0,055
L2_1_09	-0,836	0,055	-0,727	0,018	-0,825	-0,011	0,058
L2_1_12	-0,304	0,052	-0,142	0,017	-0,267	-0,037	0,055
L2_1_13	0,328	0,053	0,507	0,017	0,351	-0,024	0,056
L2_1_18	-0,375	0,053	-0,296	0,017	-0,414	0,039	0,056
L2_1_20	0,032	0,052	0,082	0,017	-0,054	0,085	0,055
L2_2_02	-1,093	0,057	-1,033	0,019	-1,117	0,023	0,060
L2_2_03	-0,035	0,052	0,135	0,017	-0,003	-0,031	0,055
L2_2_05	-0,494	0,053	-0,441	0,017	-0,552	0,058	0,056
L2_2_06	-0,829	0,055	-0,771	0,018	-0,867	0,038	0,058
L2_2_07	1,542	0,061	1,702	0,019	1,491	0,051	0,064
L2_2_08	0,029	0,052	0,219	0,017	0,077	-0,047	0,055
L2_2_10	-0,082	0,052	0,133	0,017	-0,005	-0,077	0,055
L2_2_11	0,712	0,054	0,852	0,017	0,680	0,031	0,057
L2_2_12	0,442	0,053	0,585	0,017	0,426	0,017	0,056
L2_2_13	-0,021	0,052	0,119	0,017	-0,018	-0,003	0,055
L2_2_15	0,730	0,054	1,019	0,017	0,839	-0,109	0,057
L2_2_18	0,434	0,053	0,515	0,017	0,359	0,075	0,056

**Tabla 21** Constantes de equiparación para las pruebas aplicadas en la ECE

Prueba	a	b
Comprensión lectora ECE	1,108	0,592
Comprensión lectora MC	1,061	0,621
Matemática ECE	0,965	-0,050
Matemática MC	0,883	0,113
Castellano como segunda lengua	0,930	-0,121

## 5.8 NIVELES DE LOGRO Y PRESENTACIÓN DE RESULTADOS

El diseño de la ECE 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, 1999).

La determinación de los niveles de logro de la ECE se basó en las consultas realizadas para la *Evaluación Nacional 2004* (Minedu, 2005) a grupos de expertos en Educación, con la finalidad de determinar cuáles son los niveles de logro pertinentes para clasificar a los estudiantes según su desempeño en las pruebas. Después, estos criterios fueron revisados por los especialistas de la UMC en función a las características específicas de la ECE 2007<sup>8</sup> y considerando el valor del índice de separación de personas. A partir de ello, se llegó a la conclusión de que podían diferenciarse adecuadamente tres grupos de logro.

Para dicha labor, se partió del análisis de cada una de las preguntas que formaron parte de la prueba. Estas preguntas se ordenaron, de acuerdo con su nivel de dificultad, desde la más fácil hasta la más difícil. Los especialistas establecieron hasta qué pregunta por lo menos —de la escala ordenada por dificultad— tendría que responder un estudiante para poder ser incluido en cada uno de los niveles de desempeño definidos. Para ello se usaron los datos de la ECE 2007, que estaban centrados en 0, considerando la media aritmética de la medida de dificultad de los ítems; además usaban un tamaño de *logit* igual a 1. A partir de dicho proceso, se identificaron los ítems y sus respectivas medidas de dificultad que marcaban dos puntos de corte.

Como se presentó con anterioridad, el modelo Rasch para ítems dicotómicos 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 ( $P_{ni}$ ), aplicando la ecuación 4.5.

Considerando lo trabajado en la *Evaluación Nacional 2004* (Minedu, 2005), se considera 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 a un ítem implica una medida de habilidad que sea medio *logit* más que la medida de dificultad de dicho ítem.

En las pruebas ECE, al haber dos puntos de corte, se estructuran tres niveles de menor a mayor logro, diferenciando a los estudiantes en tres grupos según los criterios que aparecen a continuación:

- aquellos estudiantes con una probabilidad menor a 0,62 de responder un ítem con una dificultad igual a la del primer corte;
- aquellos con una probabilidad mayor o igual de responder un ítem con una dificultad igual a la del primer corte y una probabilidad menor a 0,62 de responder un ítem con una dificultad igual a la del segundo corte;
- aquellos estudiantes con una probabilidad mayor o igual a 0,62 de responder un ítem con una dificultad igual a la del segundo corte.

<sup>8</sup> En el caso de Castellano como segunda lengua, se reevaluó el contenido de las pruebas, el ordenamiento de los ítems y se estableció nuevos puntos de corte en el año 2012.

Se presenta, a continuación, la medida de dificultad que marca cada punto de corte y la medida de habilidad de una persona que tenga una probabilidad igual a 0,62 de responder correctamente un ítem con dicha dificultad:

**Tabla 22** Medidas de dificultad de los ítems que marcan los puntos de corte y habilidad de una persona requerida para responderlo correctamente con P (0,62).

Corte	Comprensión de textos		Matemática		Castellano como segunda lengua	
	Dificultad	Habilidad	Dificultad	Habilidad	Dificultad	Habilidad
Nivel 1 vs. Nivel 2	2,017	2,517	1,859	2,359	1,044	1,544
Nivel 1 vs. Nivel 1	0,047	0,547	0,106	0,606	-0,016	0,484

Finalmente, es importante señalar que las medidas de habilidad de las personas fueron transformadas linealmente ( $Y = a + bX$ ), de tal manera que en la ECE 2007 su media aritmética fue 500 y su desviación estándar 100. Para ello, se usaron las constantes que aparecen a continuación:

**Tabla 23** Valores de a y b utilizados para la transformación lineal de las medidas de las pruebas ECE

Prueba	a	b
Comprensión lectora	423,614	63,585
Matemática	468,321	72,440
Castellano como segunda lengua	472,600	85,600



## 6. Referencias

- Aiken, L. (1996). *Tests psicológicos y evaluación* (8ª ed.). México: Prentice Hall.
- American Educational Research Association (AERA), American Psychological Association (APA), y National Council of Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Bond, T.G. y Fox, C.M. (2007). *Applying the Rasch model: fundamental measurement in the human Sciences* (2ª 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.
- Charter, R. (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.
- Charter, R.A. y Feldt, L.S. (2002). The importance of reliability as it relates to true score confidence intervals. *Measurement and evaluation in counseling and development*, 35(2), 104-112.
- Christensen, K. B. y Kreiner, S. (2013). Item fit statistics. En K. B. Christensen, S. Kreiner y M. Mesbah (Eds.), *Rasch models in health* (pp. 83-103). Londres: ISTE. Cochran, W. G. (1977). *Sampling Techniques*. New York, NY: John Wiley y Sons.
- Cronbach, L. y Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cuesta, M. (1996). Unidimensionalidad. En J. Muñoz (Ed.), *Psicometría* (pp. 239-291). Madrid: Pirámide.
- Embretson, S. y Reise, S. (2000). *Item Response Theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Escurra, M. (1998). Cuantificación de la validez de contenido por criterio de jueces. *Revista de Psicología*, 6(1-2), 103-111.
- Fisher, W. (1992). Reliability statistics. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 2*, 1996 (p. 238). Chicago: MESA Press.
- Fox, C. (1999). An introduction to the partial credit model for developing nursing assessments. *Journal of Nursing Education*, 38(8), 340-346.
- George, L.A., Schkade, J. K. y Ishee, J.H. (2004). Content validity of the relative mastery measurement scale: a measure of occupational adaptation. *OTJR: Occupation, Participation and Health*, 24(3), 92-102.
- Haertel, E.H. (2006). Reliability. En R.L. Brennan (editor). *Educational measurement* (pp. 17-64). Wesport, CT: American Council on Education y Praeger Publishers.
- 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.
- Head, B.J., Aquilino, M.L., Johnson, M. Reed, D., Maas, M. y Moorhead, S. (2004). Content validity and nursing sensitivity of community-level outcomes from the Nursing Outcomes Classification (NOC). *Journal of Nursing Scholarship*, 36(3), 251-259.
- Ingebo, G. (1997). *Probability in the measure of achievement*. Chicago: MESA.
- Instituto Nacional de Estadística e Informática [INEI] (2007). *Censo de población y vivienda (CPV) 2007*. Recuperado de <http://www.inei.gov.pe>.
- Kane, M. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342.
- Kane, M. (2006). Validation. En R.L. Brennan (Ed.), *Educational measurement* (4ª ed., pp. 17-64). Wesport, CT: American Council on Education y Praeger Publishers.
- Kelley, T., Ebel, R., y Linacre, J.M. (2002). Item discrimination indices. *Rasch Measurement Transactions*, 16(3), pp. 883-884. Recuperado de <http://www.rasch.org/rmt/rmt163a.htm>
- Kish, L. (1965). *Survey Sampling*. Michigan: John Wiley y Sons.
- Kolen, M. y Brennan, R. (2004). *Test equating, scaling and linking*. (2ª ed.). Nueva York, NY: Springer.
- Kubinger, K. (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology Science*, 45(1), 106-110.
- Lehtonen, R. y Pahkinen, E. (2004). *Practical Methods for Design and Analysis of Complex Surveys* (2ª ed.). New York: John Wiley y Sons.
- Linacre J. M. (1998). Detecting multidimensionality: which residual data-type works best? *Journal of Outcome Measurement*, 2(3), 266-283.
- Linacre, J. M. (2005). *Winsteps manual*. Chicago.
- Linacre, J. M. (2013). *Winsteps* (Versión 3.80.1) [Programa informático]. Beaverton, Oregon:
- Linacre, J. M. y Wright, B. D. (1989). Length of a Logit. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1*, 1995 (pp. 54-55). Chicago: MESA Press.
- Linn, R.L. (2006). The standard for educational and psychological testing: guidance in test development. En S.M. Downing y T.M. Haladyna (Eds.), *Handbook of test development* (pp. 17-28). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Livinstone, S. (2004). *Equating Test Scores (Without IRT)*. Educational Testing Service. Recuperado de <http://www.ets.org/Media/Research/pdf/LIVINGSTON.pdf>.
- Luppescu, S. (1991). Graphical diagnosis. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1*, 1995 (p. 136). Chicago: MESA Press.
- McGartland, D., Berg-Weger, M., Tebb, S.S., Lee, E.S. y Rauch, S. (2003). Objectifying content validity: conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104.
- Ministerio de Educación del Perú (2005). *Evaluación Nacional del rendimiento estudiantil 2004. Informe descriptivo de resultados*. Lima: Ministerio de Educación del Perú.
- Ministerio de Educación del Perú (2009). *Evaluación Censal de Estudiantes (ECE) Segundo grado de primaria y cuarto grado de primaria de IE EIB. Marco de trabajo*. Descargado del sitio web de la Unidad de Medición de la Calidad Educativa del Ministerio de Educación del Perú: [http://www2.minedu.gob.pe/umc/ece/Marco\\_de\\_Trabajo\\_ECE.pdf](http://www2.minedu.gob.pe/umc/ece/Marco_de_Trabajo_ECE.pdf).
- Ministerio de Educación del Perú [MINEDU] (2013). *Estadística de la calidad educativa (Escale) 2013*. Recuperado de <http://escale.minedu.gob.pe>.

- Muñiz, J. (1999). *Teoría clásica de los tests*, (2ª ed.). Madrid: Ediciones Pirámide.
- Navas, M. (1996). Equiparación de puntuaciones. En J. Muñiz (Ed.), *Psicometría* (pp. 293-370). Madrid: Universitas.
- Nunnally, J. y Bernstein, I. (1995). *Teoría psicométrica* (3ª 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.
- Paz, M. (1996). Validez. En J. Muñiz (Ed.), *Psicometría* (pp. 293-370). Madrid: Universitas.
- Perkins A. y Engelhard, G. Jr. (2009). Crossing person response functions. *Rasch Measurement Transactions*, 23 (1), pp. 1183-1184. Descargado del sitio web del Institute for Objective Measurement, <http://www.rasch.org/rmt/rmt231a.htm>
- Preacher, K.J y Robert C. MacCallum, R.C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2(1), 13-43.
- Prieto, G. y Delgado, A. R. (1996). Construcción de ítems. En J. Muñiz (Ed.), *Psicometría* (pp. 105-138). Madrid: Universitas.
- Sawilowsky, S.S. (2003). Reliability as psychometrics versus datametrics. En B. Thomson (Ed.), *Score reliability. Comotemporary thinking on reliability issues* (pp.103-122). Thousand Oaks, CA: Sage Publications.
- Schmeiser, C. y Welch, C. (2006). Test development. En R. Brennan (Ed.), *Educational measurement* (4ª ed., pp. 307-353). Wesport: Praeger.
- Schumacker, R. (2007). Reliability: A Rasch perspective. *Educational and Psychological Measurement*, 67(3), 394-409.
- Shaw, F. (1991). Descriptive IRT vs. prescriptive Rasch. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1*, 1995 (p. 131). Chicago: MESA Press.
- 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: MESA Press.
- Stevens, S.S. (1951). Matemáticas, measurmente and psychophysics. En S.S. Stevens (Ed.), *Handbook of experimental psychology*, (pp. 1-49). Oxford: Wiley.
- Suen, H. (1990). *The Psychometric Process*. Pennsylvania State University.
- Thomson, B. (2003). Understanding reliability and coefficient alpha, really. En B. Thomson (Ed.), *Score reliability. Contemporary thinking on reliability issues* (pp.3-30). Thousand Oaks, CA: Sage Publications.
- Thomson, B. y Vacha-Haase, T. (2003). Psychometrics is datametrics. The test is not reliable. En B. Thomson (Ed.), *Score reliability. Contemporary thinking on reliability issues* (pp.123-148). Thousand Oaks, CA: Sage Publications.
- Waller, N (2001). *MicroFACT 2.0. A Microcomputer Factor Analysis Program for Ordered Polytomous Data and Mainframe Size Problems*. Saint Paul: Assessment Systems Corporation.
- Wilson, M. (2005). *Constructing measures. An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Woodruff, R. S. (1971). A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association*, 66(334), 411-414.
- Wright B. D. (1991). Scores, reliabilities and assumptions. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1*, 1995 (pp. 157-158). Chicago: MESA Press.
- Wright, B. D. (1988). Some comments about guessing. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1*, 1995 (p. 9). Chicago: MESA Press.
- Wright, B. D. (1989). Rasch model from counting right answers. En J. Linacre (Ed.), *Rasch Measurement Transactions Part 1*, 1995 (p. 62). Chicago: MESA Press.
- Wright, B. D. (1990). What is information? En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1*, 1995 (p. 109). Chicago: MESA Press.
- Wright, B. D. y Linacre, J. M. (1987). Dichotomous Rasch model derived from objectivity. En J. M. Linacre (Ed.), *Rasch Measurement Transactions Part 1*, 1995 (pp. 5-6). Chicago: MESA Press.
- 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: MESA Press.
- Wright, B. D. y Stone, M. (1998). *Diseño de mejores pruebas*. México: CENEVAL.
- Wright, B. D. y Stone, M. (1999). *Measurement Essential*, 2.ª ed. Delaware: Wide Range.
- Wright, B. D. y Masters, G. (1982). *Rating Scale Analysis*. Chicago: MESA.
- Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise and Sport*, 69(1), 11-23.

Anexos

# Anexo 1

## Tablas de especificaciones de comprensión lectora. ECE 2013 segundo grado

**Tabla 1** Distribución de ítems por capacidad. Comprensión Lectora ECE 2013

Capacidad	Total	Porcentaje de la prueba
Lee oraciones	2	4,3%
Localiza información literal	17	37,0%
Infiere información	27	58,7%
<b>TOTAL</b>	<b>46</b>	<b>100%</b>

**Tabla 2** Distribución de textos en la prueba. Comprensión Lectora ECE 2013

Tipo de texto	Categoría/ Género textual	Nombre del texto	Total de ítems por texto	Total de ítems por tipo textual	% de ítems por tipo textual
Oraciones	Oración	Carla nada en el río y Víctor lee en el bote.	1	4	8,7%
		Micaela limpia la cocina y Jerónimo lava los platos.	1		
		Felipe corta una manzana y Cristina come helado.	1		
		María pinta con plumones y Alonso lee un libro de cuentos.	1		
Narrativo	Anécdota	Vicente	3	18	39,1%
		Leonardo	3		
	Cuento	Tres otorongos	6		
		Jaime y Eduardo	6		
Informativo	Descripción enciclopédica	Tarwi	4	13	28,3%
	Nota	Tía Doris	3		
	Carta de amistad	Josefina	2		
	Artículo enciclopédico	Alimentos dañinos	2		
		Animales de transporte	2		
Argumentativo	Aviso/Afiche	Concurso escolar de dibujo	3	6	13,0%
	Diálogo	Televisión	3		
Instructivo	Recomendaciones	Haz crecer una planta	5	5	10,9%

**Tabla 3** Distribución de ítems por indicador, según cuadernillo. Comprensión Lectora ECE 2013

Orden	Capacidad	Indicador	Contenido	Cuadernillo
01	Lee oraciones.	Establece correspondencia entre una oración y un dibujo.	Oración dibujo	1
02	Localiza información literal.	Identifica datos explícitos.	Oración coordinada	1
03	Localiza información literal.	Identifica datos explícitos.	Narrativo/Anécdota	1
04	Infiere información.	Deduca relaciones de causa-efecto.	Narrativo/Anécdota	1
05	Infiere información.	Deduca el tema central del texto.	Narrativo/Anécdota	1
06	Localiza información literal.	Identifica datos explícitos.	Informativo/Nota	1
07	Infiere información.	Deduca relaciones de causa-efecto.	Informativo/Nota	1
08	Infiere información.	Deduca el propósito del texto.	Informativo/Nota	1
09	Localiza información literal.	Identifica datos explícitos.	Informativo/Párrafo inductivo	1
10	Infiere información.	Deduca la idea principal de un párrafo del texto.	Informativo/Párrafo inductivo	1
11	Localiza información literal.	Identifica datos explícitos.	Argumentativo/Aviso	1
12	Localiza información literal.	Identifica datos explícitos.	Argumentativo/Aviso	1
13	Infiere información.	Deduca el propósito del texto.	Argumentativo/Aviso	1
14	Localiza información literal.	Reconoce sucesiones de hechos o acciones.	Narrativo/Cuento	1
15	Infiere información.	Deduca relaciones de causa-efecto.	Narrativo/Cuento	1
16	Infiere información.	Deduca el significado de palabras o expresiones a partir del contexto.	Narrativo/Cuento	1
17	Infiere información.	Deduca las cualidades o defectos de los personajes de una narración.	Narrativo/Cuento	1
18	Infiere información.	Deduca el tema central del texto.	Narrativo/Cuento	1
19	Infiere información.	Deduca la enseñanza del texto.	Narrativo/Cuento	1
20	Localiza información literal.	Identifica datos explícitos.	Informativo/Descripción	1
21	Infiere información.	Compara información presente en un texto.	Informativo/Descripción	1
22	Infiere información.	Deduca el significado de palabras o expresiones a partir del contexto.	Informativo/Descripción	1
23	Infiere información.	Deduca el tema central del texto.	Informativo/Descripción	1
24	Lee oraciones.	Establece correspondencia entre una oración y un dibujo.	Oración dibujo	2
25	Localiza información literal.	Identifica datos explícitos.	Oración coordinada	2
26	Localiza información literal.	Identifica datos explícitos.	Narrativo/Anécdota	2
27	Infiere información.	Deduca relaciones de causa-efecto.	Narrativo/Anécdota	2
28	Infiere información.	Deduca el tema central del texto.	Narrativo/Anécdota	2
29	Localiza información literal.	Identifica datos explícitos.	Informativo/Carta	2
30	Infiere información.	Deduca el tema central del texto.	Informativo/Carta	2
31	Localiza información literal.	Identifica datos explícitos.	Informativo/Párrafo inductivo	2
32	Infiere información.	Deduca la idea principal de un párrafo del texto.	Informativo/Párrafo inductivo	2
33	Infiere información.	Compara información presente en un texto.	Argumentativo/Diálogo	2
34	Localiza información literal.	Identifica datos explícitos.	Argumentativo/Diálogo	2
35	Infiere información.	Compara información presente en un texto.	Argumentativo/Diálogo	2
36	Localiza información literal.	Reconoce sucesiones de hechos o acciones.	Narrativo/Cuento	2
37	Infiere información.	Deduca relaciones causa-efecto.	Narrativo/Cuento	2
38	Infiere información.	Deduca el significado de palabras o expresiones a partir del contexto.	Narrativo/Cuento	2
39	Infiere información.	Deduca las cualidades o defectos de los personajes de una narración.	Narrativo/Cuento	2
40	Infiere información.	Deduca la enseñanza del texto.	Narrativo/Cuento	2
41	Infiere información.	Deduca el tema central del texto.	Narrativo/Cuento	2
42	Localiza información literal.	Reconoce sucesiones de hechos o acciones.	Instructivo/Recomendaciones	2
43	Localiza información literal.	Identifica datos explícitos.	Instructivo/Recomendaciones	2
44	Localiza información literal.	Identifica datos explícitos.	Instructivo/Recomendaciones	2
45	Infiere información.	Deduca relaciones causa-efecto.	Instructivo/Recomendaciones	2
46	Infiere información.	Deduca el propósito del texto.	Instructivo/Recomendaciones	2

## Anexo 2

Tablas de especificaciones de comprensión lectora. MC-2013  
segundo grado**Tabla 4** Distribución de ítems por capacidad. Comprensión Lectora MC-2013

Capacidad	Total	Porcentaje de la prueba
Lee oraciones	2	4,4%
Localiza información literal	14	30,4%
Infiere información	30	65,2%
<b>TOTAL</b>	<b>46</b>	<b>100%</b>

**Tabla 5** Distribución de textos en la prueba. Comprensión Lectora MC-2013

Tipo de texto	Categoría/ Género textual	Nombre del texto	Total de ítems por texto	Total de ítems por tipo textual	% de ítems por tipo textual
Oraciones	Oración	Carla nada en el río y Víctor lee en el bote.	1	3	6,5%
		Micaela limpia la cocina y Jerónimo lava los platos.	1		
		La mamá lava la ropa y el papá plancha un pantalón.	1		
Narrativo	Anécdota	Vicente	3	18	39,1%
		El dibujo de Martín	3		
	Cuento	Tres otorongos	6		
		Campesino	6		
Informativo	Descripción enciclopédica	Tarwi	4	22	47,8%
	Descripción de personas	Jonás	5		
	Nota	Tía Doris	3		
		Leticia	3		
	Artículo enciclopédico	Alimentos dañinos	2		
		La vaca	2		
Cuadro comparativo	Chompas	3			
Argumentativo	Aviso/Afiche	Concurso escolar de dibujo	3	3	6,5%

**Tabla 6** Distribución de ítems por indicador, según cuadernillo. Comprensión Lectora MC-2013

Orden	Capacidad	Indicador	Contenido	Cuadernillo
01	Lee oraciones.	Establece correspondencia entre una oración y un dibujo.	Oración dibujo	1
02	Localiza información literal.	Identifica datos explícitos.	Oración coordinada	1
03	Localiza información literal.	Identifica datos explícitos.	Narrativo/Anécdota	1
04	Infiere información.	Deduca relaciones de causa-efecto.	Narrativo/Anécdota	1
05	Infiere información.	Deduca el tema central del texto.	Narrativo/Anécdota	1
06	Localiza información literal.	Identifica datos explícitos.	Informativo/Nota	1
07	Infiere información.	Deduca relaciones de causa-efecto.	Informativo/Nota	1
08	Infiere información.	Deduca el propósito del texto.	Informativo/Nota	1
09	Localiza información literal.	Identifica datos explícitos.	Informativo/ Párrafo inductivo	1
10	Infiere información.	Deduca la idea principal de un párrafo del texto.	Informativo/Párrafo inductivo	1
11	Localiza información literal.	Identifica datos explícitos.	Argumentativo/Aviso	1
12	Localiza información literal.	Identifica datos explícitos.	Argumentativo/Aviso	1
13	Infiere información.	Deduca el propósito del texto.	Argumentativo/Aviso	1
14	Localiza información literal.	Reconoce sucesiones de hechos o acciones.	Narrativo/Cuento	1
15	Infiere información.	Deduca relaciones de causa-efecto.	Narrativo/Cuento	1
16	Infiere información.	Deduca el significado de palabras o expresiones a partir del contexto.	Narrativo/Cuento	1
17	Infiere información.	Deduca las cualidades o defectos de los personajes de una narración.	Narrativo/Cuento	1
18	Infiere información.	Deduca el tema central del texto.	Narrativo/Cuento	1
19	Infiere información.	Deduca la enseñanza del texto.	Narrativo/Cuento	1
20	Localiza información literal.	Identifica datos explícitos.	Informativo/Descripción	1
21	Infiere información.	Compara información presente en un texto.	Informativo/Descripción	1
22	Infiere información.	Deduca el significado de palabras o expresiones a partir del contexto.	Informativo/Descripción	1
23	Infiere información.	Deduca el tema central del texto.	Informativo/ Descripción	1
24	Lee oraciones.	Establece correspondencia entre una oración y un dibujo.	Oración dibujo	2
25	Localiza información literal.	Identifica datos explícitos.	Informativo/Párrafo inductivo	2
26	Infiere información.	Deduca la idea principal de un párrafo del texto.	Informativo/Párrafo inductivo	2
27	Localiza información literal.	Identifica datos explícitos.	Informativo/Nota	2
28	Infiere información.	Deduca relaciones de causa-efecto.	Informativo/Nota	2
29	Infiere información.	Deduca el propósito del texto.	Informativo/Nota	2
30	Localiza información literal.	Identifica datos explícitos.	Narrativo/Anécdota	2
31	Infiere información.	Deduca relaciones de causa-efecto.	Narrativo/Anécdota	2
32	Infiere información.	Deduca el tema central del texto.	Narrativo/Anécdota	2
33	Infiere información.	Deduca relaciones de causa-efecto.	Narrativo/Cuento	2
34	Localiza información literal.	Reconoce sucesiones de hechos o acciones.	Narrativo/Cuento	2
35	Infiere información.	Deduca relaciones de causa-efecto.	Narrativo/Cuento	2
36	Infiere información.	Deduca el significado de palabras o expresiones a partir del contexto.	Narrativo/Cuento	2
37	Infiere información.	Deduca la enseñanza del texto.	Narrativo/Cuento	2
38	Infiere información.	Deduca las cualidades o defectos de los personajes de una narración.	Narrativo/Cuento	2
39	Localiza información literal.	Identifica datos explícitos.	Informativo/Cuadro comparativo	2
40	Infiere información.	Compara información presente en un texto.	Informativo/Cuadro comparativo	2
41	Infiere información.	Compara información presente en un texto.	Informativo/Cuadro comparativo	2
42	Localiza información literal.	Identifica datos explícitos.	Informativo/Descripción	2
43	Infiere información.	Deduca relaciones de causa-efecto.	Informativo/Descripción	2
44	Infiere información.	Deduca el significado de palabras o expresiones a partir del contexto.	Informativo/Descripción	2
45	Infiere información.	Deduca el tema central del texto.	Informativo/Descripción	2
46	Infiere información.	Deduca el propósito del texto.	Instructivo/Recomendaciones	2

## Anexo 3 Tablas de especificaciones de matemática ECE 2013 segundo grado

**Tabla 7** Distribución de ítems por capacidad. Matemática ECE 2013

Conocimiento	Capacidad	Razonamiento y Demostración	Comunicación Matemática	Resolución de problemas <sup>1</sup>	Total	Porcentaje de la prueba
<b>Comprensión del significado y uso de las operaciones (Adición y sustracción)</b>		--	--	27	27	64,3%
Operaciones de adición y sustracción		--	--	7	7	
Problemas asociados al significado de cambio		--	--	3	3	
Problemas asociados al significado combinación		--	--	6	6	
Problemas asociados al significado de comparación		--	--	3	3	
Problemas asociados al significado de igualación		--	--	3	3	
Problemas de doble, triple, mitad		--	--	2	2	
Problemas de varias etapas		--	--	3	3	
<b>Comprensión del significado y uso del número y del SND (Sistema de Numeración Decimal)</b>		3	7	5	15	35,7%
Comparación de números		--	2	--	2	
Secuencias numéricas		2	--	--	2	
Recodificación de representaciones numéricas		--	2	--	2	
Agrupación reiterada de grupos de 10		1	--	2	3	
Reagrupación de cantidades para identificar grupos de 10		--	--	3	3	
Representaciones y equivalencias convencionales y no convencionales		--	3	--	3	
<b>TOTAL</b>		<b>3</b>	<b>7</b>	<b>32</b>	<b>42</b>	<b>100%</b>

**Tabla 8** Distribución de ítems por indicador, según cuadernillo. Matemática ECE 2013

Orden	Capacidad	Indicador	Contenido	Cua-der-nillo
01	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la suma de dos números de dos cifras presentadas en enunciado verbal.	Comprensión del significado y uso de las operaciones	1
02	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la suma de dos números de dos cifras, presentadas en formato vertical.	Comprensión del significado y uso de las operaciones	1
03	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la diferencia de dos números de hasta dos cifras, presentadas en enunciado verbal.	Comprensión del significado y uso de las operaciones	1
04	Razonamiento y demostración	Identifica la secuencia numérica que cumple con un patrón dado.	Comprensión del significado y uso del número y del SND	1
05	Comunicación matemática	Señala los números mayores o menores respecto de un referente.	Comprensión del significado y uso del número y del SND	1
06	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "juntar" con información presentada en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
07	Resolución de problemas	Resuelve situaciones aditivas de varias etapas, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
08	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "igualar" con información presentada en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
09	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "juntar" con información presentada en tablas de doble entrada.	Comprensión del significado y uso de las operaciones	1
10	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de juntar cantidades y formar grupos de 10, sin residuo, presentadas en diversos tipos de texto.	Comprensión del significado y uso del número y del SND	1

<sup>1</sup> Se consideran los algoritmos dentro del proceso de resolución de problemas, debido a constituir uno de los procedimientos que permite resolverlos.



Orden	Capacidad	Indicador	Contenido	Cua-der-nillo
11	Resolución de problemas	Resuelve situaciones asociadas a una relación directa de doble, triple o mitad de una cantidad, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
12	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "comparar", con contradicción en la información, a partir de diferentes tipos de texto.	Comprensión del significado y uso de las operaciones	1
13	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "separar" a partir de información presentada en texto continuo.	Comprensión del significado y uso de las operaciones	1
14	Comunicación matemática	Expresa números menores que 100 en su representación compacta usual desde su representación gráfica.	Comprensión del significado y uso del número y del SND	1
15	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "quitar" a partir de información presentada en texto continuo.	Comprensión del significado y uso de las operaciones	1
16	Comunicación matemática	Expresa números menores que 100 desde su representación gráfica a su notación expresada en decenas.	Comprensión del significado y uso del número y del SND	1
17	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de juntar cantidades y formar grupos de 10, con residuo, presentadas en diversos tipos de texto.	Comprensión del significado y uso del número y del SND	1
18	Resolución de problemas	Resuelve situaciones aditivas en acciones de "comparar" a partir de información presentada en texto continuo.	Comprensión del significado y uso de las operaciones	1
19	Resolución de problemas	Resuelve situaciones aditivas de varias etapas, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
20	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "igualar" a partir de información presentada en diversos tipos de texto	Comprensión del significado y uso de las operaciones	1
21	Resolución de problemas	Identifica la agrupación reiterada de 10 unidades a partir de un enunciado de texto.	Comprensión del significado y uso del número y del SND	1
22	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la suma de dos sumandos de dos cifras, presentadas en formato vertical.	Comprensión del significado y uso de las operaciones	2
23	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la diferencia de dos números de dos cifras, presentadas en enunciado verbal.	Comprensión del significado y uso de las operaciones	2
24	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la diferencia de dos números de dos cifras, presentadas en formato vertical.	Comprensión del significado y uso de las operaciones	2
25	Resolución de problemas	Resuelve situaciones aditivas referidas a la suma de dos números de dos cifras, donde se pide hallar uno de los sumandos.	Comprensión del significado y uso de las operaciones	2
26	Razonamiento y demostración	Halla el patrón de una secuencia numérica sencilla para completar el término que falta.	Comprensión del significado y uso del número y del SND	2
27	Comunicación matemática	Señala el número mayor o menor entre tres cantidades.	Comprensión del significado y uso del número y del SND	2
28	Resolución de problemas	Resuelve situaciones aditivas en acciones de "igualar", presentadas con soporte gráfico	Comprensión del significado y uso de las operaciones	2
29	Comunicación matemática	Expresa números menores que 100 desde su representación gráfica a su notación expresada en decenas.	Comprensión del significado y uso del número y del SND	2
30	Comunicación matemática	Expresa la representación no convencional en unidades de orden en números de dos cifras.	Comprensión del significado y uso del número y del SND	2
31	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "juntar" a partir de información presentada en un soporte gráfico.	Comprensión del significado y uso de las operaciones	2
32	Comunicación matemática	Expresa números menores que 100 desde su representación gráfica a su representación compacta usual.	Comprensión del significado y uso del número y del SND	2
33	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "quitar" en la que se pide hallar la cantidad que produce el cambio, presentadas en texto continuo.	Comprensión del significado y uso de las operaciones	2
34	Resolución de problemas	Resuelve situaciones asociadas a la relación directa de doble, triple o mitad de una cantidad, presentada en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	2
35	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "juntar" a partir de información presentada en tablas de doble entrada.	Comprensión del significado y uso de las operaciones	2
36	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de juntar cantidades y formar grupos de 10, sin residuo, presentadas en diversos tipos de texto.	Comprensión del significado y uso del número y del SND	2
37	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "quitar", presentadas en texto continuo y con información adicional.	Comprensión del significado y uso de las operaciones	2
38	Razonamiento y demostración	Identifica la agrupación reiterada de 10 unidades a partir de información presentada en diversos tipos de texto.	Comprensión del significado y uso del número y del SND	2
39	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "separar" a partir de información presentada en enunciado verbal.	Comprensión del significado y uso de las operaciones	2
40	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de comparar, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	2
41	Resolución de problemas	Resuelve situaciones vinculadas a la agrupación reiterada de 10 unidades donde debe encontrar la cantidad que le falta para completar una decena más.	Comprensión del significado y uso del número y del SND	2
42	Resolución de problemas	Resuelve situaciones aditivas de varias etapas, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	2

## Anexo 4 Tablas de especificaciones de matemática MC-2013. segundo grado

**Tabla 9** Distribución de ítems por capacidad. Matemática MC-2013

Conocimiento	Capacidad	Razonamiento y Demostración	Comunicación Matemática	Resolución de problemas <sup>2</sup>	Total	Porcentaje de la prueba
<b>Comprensión del significado y uso de las operaciones (Adición y sustracción)</b>		--	--	27	27	64,3%
Operaciones de adición y sustracción		--	--	7	7	
Problemas asociados al significado de cambio		--	--	2	2	
Problemas asociados al significado combinación		--	--	7	7	
Problemas asociados al significado de comparación		--	--	2	2	
Problemas asociados al significado de igualación		--	--	3	3	
Problemas de doble, triple, mitad		--	--	2	2	
Problemas de varias etapas		--	--	4	4	
<b>Comprensión del significado y uso del número y del SND (Sistema de Numeración Decimal)</b>		2	8	5	15	35,7%
Comparación de números		--	2	--	2	
Secuencias numéricas		2	--	--	2	
Recodificación de representaciones numéricas		--	1	--	1	
Agrupación reiterada de grupos de 10		--	--	2	2	
Reagrupación de cantidades para identificar grupos de 10		--	--	3	3	
Representaciones y equivalencias convencionales y no convencionales		--	5	--	5	
<b>TOTAL</b>		<b>2</b>	<b>8</b>	<b>32</b>	<b>42</b>	<b>100%</b>

**Tabla 10** Distribución de ítems por indicador, según cuadernillo. Matemática MC-2013

Orden	Capacidad	Indicador	Contenido	Cuadernillo
01	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la suma de dos números de dos cifras presentadas en enunciado verbal.	Comprensión del significado y uso de las operaciones	1
02	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la suma de dos números de dos cifras, presentadas en formato vertical.	Comprensión del significado y uso de las operaciones	1
03	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la diferencia de dos números de hasta dos cifras, presentadas en enunciado verbal.	Comprensión del significado y uso de las operaciones	1
04	Razonamiento y demostración	Identifica la secuencia numérica que cumple con un patrón dado.	Comprensión del significado y uso del número y del SND	1
05	Comunicación matemática	Señala los números mayores o menores respecto de un referente.	Comprensión del significado y uso del número y del SND	1
06	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "juntar" con información presentada en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
07	Resolución de problemas	Resuelve situaciones aditivas de varias etapas, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
08	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "igualar" con información presentada en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
09	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "juntar" con información presentada en tablas de doble entrada.	Comprensión del significado y uso de las operaciones	1
10	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de juntar cantidades y formar grupos de 10, sin residuo, presentadas en diversos tipos de texto.	Comprensión del significado y uso del número y del SND	1

<sup>2</sup>Se consideran los algoritmos dentro del proceso de resolución de problemas, debido a constituir uno de los procedimientos que permite resolverlos.

Orden	Capacidad	Indicador	Contenido	Cua- derno
11	Resolución de problemas	Resuelve situaciones asociadas a una relación directa de doble, triple o mitad de una cantidad, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
12	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "comparar", con contradicción en la información, a partir de diferentes tipos de texto.	Comprensión del significado y uso de las operaciones	1
13	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "separar" a partir de información presentada en texto continuo.	Comprensión del significado y uso de las operaciones	1
14	Comunicación matemática	Expresa números menores que 100 en su representación compacta usual desde su representación gráfica.	Comprensión del significado y uso del número y del SND	1
15	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "quitar" a partir de información presentada en texto continuo.	Comprensión del significado y uso de las operaciones	1
16	Comunicación matemática	Expresa números menores que 100 desde su representación gráfica a su notación expresada en decenas.	Comprensión del significado y uso del número y del SND	1
17	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de juntar cantidades y formar grupos de 10, con residuo, presentadas en diversos tipos de texto.	Comprensión del significado y uso del número y del SND	1
18	Resolución de problemas	Resuelve situaciones aditivas en acciones de "comparar" a partir de información presentada en texto continuo.	Comprensión del significado y uso de las operaciones	1
19	Resolución de problemas	Resuelve situaciones aditivas de varias etapas, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	1
20	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "igualar" a partir de información presentada en diversos tipos de texto	Comprensión del significado y uso de las operaciones	1
21	Resolución de problemas	Identifica la agrupación reiterada de 10 unidades a partir de un enunciado de texto.	Comprensión del significado y uso del número y del SND	1
22	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la suma de dos sumandos de dos cifras, presentadas en enunciado verbal.	Comprensión del significado y uso de las operaciones	2
23	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la diferencia de dos números de dos cifras, presentadas en formato vertical.	Comprensión del significado y uso de las operaciones	2
24	Resolución de problemas	Resuelve situaciones aditivas donde se pide hallar la suma de dos sumandos de dos cifras, presentadas en formato vertical.	Comprensión del significado y uso de las operaciones	2
25	Resolución de problemas	Resuelve situaciones aditivas referidas a la suma de cinco sumandos de hasta dos cifras, presentadas en formato horizontal.	Comprensión del significado y uso de las operaciones	2
26	Comunicación matemática	Señala el número mayor o menor entre tres cantidades.	Comprensión del significado y uso del número y del SND	2
27	Razonamiento y demostración	Halla el patrón de una secuencia numérica sencilla para completar el término que falta.	Comprensión del significado y uso del número y del SND	2
28	Resolución de problemas	Resuelve situaciones aditivas en acciones de "igualar", presentadas con soporte gráfico	Comprensión del significado y uso de las operaciones	2
29	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "quitar" a partir de información presentada en texto continuo.	Comprensión del significado y uso de las operaciones	2
30	Resolución de problemas	Resuelve situaciones aditivas de varias etapas, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	2
31	Comunicación matemática	Expresa números menores que 100 desde su representación gráfica a su notación expresada en decenas.	Comprensión del significado y uso del número y del SND	2
32	Comunicación matemática	Expresa la representación no convencional en unidades de orden en números de dos cifras.	Comprensión del significado y uso del número y del SND	2
33	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "juntar" a partir de información presentada en un soporte gráfico.	Comprensión del significado y uso de las operaciones	2
34	Comunicación matemática	Expresa números menores que 100 desde su representación gráfica a su notación expresada en decenas.	Comprensión del significado y uso del número y del SND	2
35	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "separar" a partir de información presentada en enunciado verbal.	Comprensión del significado y uso de las operaciones	2
36	Resolución de problemas	Resuelve situaciones asociadas a la relación directa de doble, triple o mitad de una cantidad, presentada en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	2
37	Comunicación matemática	Expresa la representación no convencional en unidades de orden en números de dos cifras.	Comprensión del significado y uso del número y del SND	2
38	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de juntar cantidades y formar grupos de 10, con residuo, presentadas en diversos tipos de texto.	Comprensión del significado y uso del número y del SND	2
39	Resolución de problemas	Resuelve situaciones aditivas asociadas a acciones de "juntar" a partir de información presentada en tablas de doble entrada.	Comprensión del significado y uso de las operaciones	2
40	Comunicación matemática	Expresa equivalencias entre unidades de orden en números de dos cifras.	Comprensión del significado y uso del número y del SND	2
41	Razonamiento y demostración	Identifica la agrupación reiterada de 10 unidades a partir de información presentada en diversos tipos de texto.	Comprensión del significado y uso del número y del SND	2
42	Resolución de problemas	Resuelve situaciones aditivas de varias etapas, presentadas en diversos tipos de texto.	Comprensión del significado y uso de las operaciones	2

## Anexo 5

## Tablas de especificaciones de comprensión lectora - castellano como segunda lengua. ECE 2013. Cuarto grado EIB

**Tabla 11** Distribución de ítems por capacidad. Comprensión Lectora – Castellano como segunda lengua ECE 2013 cuarto grado EIB.

Capacidad	Total	Porcentaje de la prueba
Lee oraciones	2	4,6%
Localiza información literal	17	38,6%
Infiere información	25	56,8%
<b>TOTAL</b>	<b>44</b>	<b>100%</b>

**Tabla 12** Distribución de textos en la prueba. Comprensión Lectora – Castellano como segunda lengua ECE 2013 cuarto grado EIB.

Tipo de texto	Categoría/ Género textual	Nombre del texto	Total de ítems por texto	Total de ítems por tipo textual	% de ítems por tipo textual
Oraciones	Oración	Pancho y Antonia	1	6	13,6%
		Las gallinas comen gusanos	1		
		José está muy contento	1		
		Julio y Paulina	1		
		Andrés	1		
		Rolando	1		
Narrativo	Anécdota	La profesora Nilda	2	13	29,6%
	Cuento	El hombre y la culebra	6		
		La Capusa	6		
Informativo	Descripción enciclopédica	¿Sabías que los animales pueden comunicarse?	4	19	43,2%
		Conocimientos de pueblo shipibo	4		
	Nota	Liz y Arturo	2		
	Noticia	Algunas mejoras que traen daño	3		
		Interesante jornada en Plaza de Armas	5		
Cuadro comparativo	Jacinto	1			
Instructivo	Instructivo para elaborar un objeto	Las canastas de la selva	3	6	13,6%
		El café	3		

**Tabla 13** Distribución de ítems por indicador, según cuadernillo. Comprensión Lectora – Castellano como segunda lengua ECE 2013 cuarto grado EIB.

Orden	Capacidad	Indicador	Contenido	Cuadernillo
01	Lee oraciones.	Establece correspondencia entre una oración y un dibujo.	Oración dibujo	1
02	Localiza información literal.	Identifica datos explícitos.	Oración simple	1
03	Localiza información literal.	Identifica datos explícitos.	Oración subordinada	1
04	Localiza información literal.	Identifica datos explícitos.	Narrativo/Anécdota	1
05	Infiere información.	Deduce relaciones de causa-efecto.	Narrativo/Anécdota	1
06	Localiza información literal.	Identifica datos explícitos.	Informativo/Noticia	1
07	Infiere información.	Deduce el significado de palabras o expresiones a partir del contexto.	Informativo/Noticia	1
08	Infiere información.	Deduce el tema central del texto.	Informativo/Noticia	1
09	Localiza información literal.	Identifica datos explícitos.	Instructivo/Instructivo para elaborar objetos	1
10	Localiza información literal.	Reconoce sucesiones de hechos o acciones.	Instructivo/Instructivo para elaborar objetos	1
11	Infiere información.	Deduce el propósito del texto.	Instructivo/Instructivo para elaborar objetos	1
12	Infiere información.	Deduce relaciones de causa-efecto.	Informativo/Descripción	1
13	Localiza información literal.	Identifica datos explícitos.	Informativo/Descripción	1
14	Infiere información.	Deduce el tema central del texto.	Informativo/Descripción	1
15	Infiere información.	Deduce el propósito del texto.	Informativo/ Descripción	1
16	Infiere información.	Deduce relaciones de causa-efecto.	Narrativo/Cuento	1
17	Localiza información literal.	Reconoce sucesiones de hechos o acciones.	Narrativo/Cuento	1
18	Infiere información.	Deduce relaciones de causa-efecto.	Narrativo/Cuento	1
19	Infiere información.	Deduce las cualidades o defectos de los personajes de una narración.	Narrativo/Cuento	1
20	Infiere información.	Deduce el tema central del texto.	Narrativo/Cuento	1
21	Infiere información.	Deduce el significado de palabras o expresiones a partir del contexto.	Narrativo/Cuento	1
22	Infiere información.	Deduce relaciones de semejanza y diferencia	Informativo/Cuadro comparativo	1
23	Lee oraciones.	Establece correspondencia entre una oración y un dibujo.	Oración dibujo	2
24	Localiza información literal.	Identifica datos explícitos.	Oración simple	2
25	Localiza información literal.	Identifica datos explícitos.	Oración subordinada	2
26	Localiza información literal.	Identifica datos explícitos.	Informativo/Nota	2
27	Infiere información.	Deduce relaciones de causa-efecto.	Informativo/Nota	2
28	Infiere información.	Deduce relaciones de causa-efecto.	Informativo/Noticia	2
29	Localiza información literal.	Identifica datos explícitos.	Informativo/Noticia	2
3	Infiere información.	Deduce el tema central del texto.	Informativo/Noticia	2
31	Infiere información.	Deduce relaciones de causa-efecto.	Informativo/Noticia	2
32	Infiere información.	Deduce el propósito del texto.	Informativo/Noticia	2
33	Infiere información.	Deduce relaciones de causa-efecto.	Narrativo/Cuento	2
34	Localiza información literal.	Identifica datos explícitos.	Narrativo/Cuento	2
35	Infiere información.	Deduce las cualidades o defectos de los personajes de una narración.	Narrativo/Cuento	2
36	Localiza información literal.	Reconoce sucesiones de hechos o acciones.	Narrativo/Cuento	2
37	Infiere información.	Deduce el significado de palabras o expresiones a partir del contexto.	Narrativo/Cuento	2
38	Infiere información.	Deduce relaciones de causa-efecto.	Informativo/Descripción	2
39	Localiza información literal.	Identifica datos explícitos.	Informativo/Descripción	2
40	Infiere información.	Deduce el tema central del texto.	Informativo/Descripción	2
41	Infiere información.	Deduce el propósito del texto.	Informativo/Descripción	2
42	Localiza información literal.	Identifica datos explícitos.	Instructivo/Instructivo para elaborar objetos	2
43	Localiza información literal.	Reconoce sucesiones de hechos o acciones.	Instructivo/Instructivo para elaborar objetos	2
44	Infiere información.	Deduce el propósito del texto.	Instructivo/Instructivo para elaborar objetos	2

## Anexo 6

## Ejemplo del formato para el análisis de la calidad técnica de los ítems. Comprensión lectora - segundo grado primaria

Operativo: \_\_\_\_\_ N° cuadernillo: \_\_\_\_\_

Tipo textual: \_\_\_\_\_ Nombre del texto: \_\_\_\_\_

No. de preguntas: \_\_\_\_\_ Revisor: \_\_\_\_\_

Institución: \_\_\_\_\_ Fecha: \_\_\_\_\_

Escriba una “X” bajo la columna correspondiente de acuerdo con su opinión sobre las características del texto o el ítem; que se detallan a continuación:

**CONSTRUCCIÓN DEL TEXTO:**

N.	Característica del texto	SÍ	NO
1	¿El lenguaje es adecuado para niños de segundo grado? Vocabulario, registro y estructura (subordinaciones e incrustaciones adecuadas, referentes cercanos del objeto referido)		
2	¿El texto tiene algún tipo de sesgo (racial, social, cultural, de género, estereotipos)?		
3	¿Son pertinentes los recursos complementarios propuestos para el texto? (ilustraciones, tipo y tamaño de fuentes, etc.)		
4	Las preguntas planteadas a partir del texto, hacen referencia a la información más importante del texto?		

**CARACTERÍSTICAS GENERALES DE LOS ÍTEMS:**

		ÍTEM 1		ÍTEM 2		ÍTEM 3	
4.	¿Se ajusta el ítem al indicador que le corresponde?						
5.	¿La pregunta se contesta a partir de la información (explícita o implícita) del texto?						
6.	¿La respuesta correcta a este ítem no puede ser inferida a partir de otro ítem?						
7.	¿La complejidad de la tarea solicitada es adecuada para niños de segundo grado de EBR?						

**CONSTRUCCIÓN DEL ENUNCIADO DEL ÍTEM**

		ÍTEM 1		ÍTEM 2		ÍTEM 3	
8.	¿El enunciado del ítem está claramente redactado? (vocabulario, estructura gramatical, etc.)						
9.	¿Está el enunciado libre de elementos irrelevantes (ilustraciones, textos, etc.)?						

**CARACTERÍSTICAS GENERALES DE LOS ÍTEMS:**

		ÍTEM 1		ÍTEM 2		ÍTEM 3	
10.	¿Hay una sola respuesta correcta?						
11.	¿Se han empleado los posibles errores frecuentes del niño para elaborar los distractores?						
12.	¿Están las alternativas libres de palabras, frases o imágenes irrelevantes?						

Se sugiere las siguientes revisiones:

---



---



---

## Anexo 7

## Ejemplo del formato para el análisis de la calidad técnica de los ítems. Matemática - segundo grado primaria

Operativo: \_\_\_\_\_ N° cuadernillo: \_\_\_\_\_

No. de preguntas \_\_\_\_\_ Fecha: \_\_\_\_\_

Revisor: \_\_\_\_\_ Institución: \_\_\_\_\_

Escriba una "X" bajo la columna correspondiente de acuerdo con su opinión sobre las características del texto o el ítem; que se detallan a continuación:

			ÍTEM 1		ÍTEM 2		ÍTEM 3	
	N°	Característica	SI	NO	SI	NO	SI	NO
CARACTERÍSTICAS GENERALES	1.	¿Se ajusta el ítem al indicador que le corresponde?						
	2.	¿La complejidad de la tarea solicitada es adecuada para niños de segundo grado? (habilidad y contenido involucrado)						
	3.	¿Existe algún tipo de sesgo (racial, cultural, social, de sexo, estereotipo, etc.)						
	4.	¿La respuesta correcta a este ítem no puede ser inferida a partir de otro ítem?						
ANÁLISIS DEL ENUNCIADO	5.	¿El problema o tarea a realizar está claramente definida en el enunciado?						
	6.	El enunciado del ítem está claramente redactado? (estructura gramatical, vocabulario, simbología, etc.)						
	7.	¿Está el enunciado libre de elementos irrelevante? (ilustraciones, textos, etc.)						
	8.	En caso de tener ilustración, ¿ayuda a la comprensión del enunciado?						
ANÁLISIS DE LAS ALTERNATIVAS	9.	¿Hay una sola respuesta correcta?						
	10.	¿Se han empleado errores posibles en la solución del ítem como alternativas de respuesta?						
	11.	¿Están las alternativas libres de material irrelevante?						

Se sugiere las siguientes revisiones:

---



---



---



---

## Anexo 8

## Ejemplo del formato para el análisis de la calidad técnica de los ítems. Comprensión lectora en castellano como segunda lengua - cuarto grado primaria EIB

Operativo: \_\_\_\_\_ N° cuadernillo: \_\_\_\_\_

Tipo textual: \_\_\_\_\_ Nombre del texto: \_\_\_\_\_

No. de preguntas: \_\_\_\_\_ Revisor: \_\_\_\_\_

Institución: \_\_\_\_\_ Fecha: \_\_\_\_\_

Escriba una "X" bajo la columna correspondiente de acuerdo con su opinión sobre las características del texto o el ítem; que se detallan a continuación:

N.	Característica del texto	SÍ	NO
1	¿El lenguaje es adecuado para niños de cuarto grado EIB? Vocabulario y estructuras gramaticales adecuadas		
2	¿El texto tiene suficientes pistas o información para que los niños de cuarto grado EIB puedan comprender nuevas palabras o temas?		
3	¿El texto tiene algún tipo de sesgo (social, cultural, de género, racial, estereotipos)?		
4	¿Son pertinentes los recursos complementarios propuestos para el texto? (ilustraciones, tipo y tamaño de fuentes, etc.)		
5	Las preguntas planteadas a partir del texto, hacen referencia a la información más importante del texto?		

		ÍTEM 1	ÍTEM 2	ÍTEM 3
6.	¿Se ajusta el ítem al indicador que le corresponde?			
7.	¿La pregunta se contesta a partir de la información (explícita o implícita) del texto?			
8.	¿La respuesta correcta a este ítem no puede ser inferida a partir de otro ítem?			
9.	¿La complejidad de la tarea solicitada es adecuada para niños de segundo grado de EIB?			

		ÍTEM 1	ÍTEM 2	ÍTEM 3
10.	¿El enunciado del ítem está claramente redactado y presentado? (vocabulario, estructura gramatical, subrayado en la negación y adverbios de tiempo, etc.)			
11.	¿Está el enunciado libre de elementos irrelevantes (ilustraciones, textos, etc.)?			

		ÍTEM 1	ÍTEM 2	ÍTEM 3
12.	¿Hay una sola respuesta correcta?			
13.	¿Se han empleado los posibles errores frecuentes del niño para elaborar los distractores?			
14.	¿Están las alternativas libres de palabras, frases o imágenes irrelevantes?			
15.	¿La extensión de las alternativas es adecuada y entendible para los estudiantes de cuarto grado EIB?			
16.	¿Existe concordancia gramatical de número entre las alternativas y la pregunta?			
17.	¿La alternativa tiene un sesgo de castellano dialectal?			

Se sugiere las siguientes revisiones:

---



---



## Anexo 9

## Porcentaje de selección y correlación alternativa - Medida de la prueba de comprensión lectora. ECE 2013 segundo grado

A continuación se muestran las tablas con información para cada una de las pruebas aplicadas como parte de la ECE. Es importante señalar que en la columna alternativa, "M" significa multimarca y "N" que el alumno no respondió el ítem. En la columna valor, el "1" identifica la alternativa correcta. Además, se presenta el porcentaje de selección de cada alternativa (%); así como la correlación ítem-test (ptme) ya explicada en este documento.

**Tabla 14** Características de las alternativas de la prueba de Comprensión Lectora ECE 2013

Item	Alternativa	Valor	%	ptme	Item	Alternativa	Valor	%	ptme
cl1_01	A	0	2,7	-0,17	cl1_11	A	1	70,0	0,37
cl1_01	B	0	4,2	-0,25	cl1_11	B	0	17,2	-0,24
cl1_01	C	1	91,7	0,36	cl1_11	C	0	11,0	-0,14
cl1_01	M	0	0,6	-0,10	cl1_11	M	0	0,3	-0,07
cl1_01	N	0	0,9	-0,14	cl1_11	N	0	1,5	-0,23
cl1_02	A	0	8,0	-0,25	cl1_12	A	1	76,0	0,38
cl1_02	B	1	87,7	0,39	cl1_12	B	0	17,4	-0,20
cl1_02	C	0	3,1	-0,23	cl1_12	C	0	4,7	-0,25
cl1_02	M	0	0,5	-0,09	cl1_12	M	0	0,3	-0,08
cl1_02	N	0	0,7	-0,16	cl1_12	N	0	1,7	-0,23
cl1_03	A	1	86,3	0,44	cl1_13	A	0	10,0	-0,19
cl1_03	B	0	6,8	-0,32	cl1_13	B	1	65,4	0,34
cl1_03	C	0	5,8	-0,21	cl1_13	C	0	22,3	-0,17
cl1_03	M	0	0,3	-0,08	cl1_13	M	0	0,3	-0,08
cl1_03	N	0	0,9	-0,20	cl1_13	N	0	2,0	-0,22
cl1_04	A	0	3,0	-0,23	cl1_14	A	0	18,2	-0,23
cl1_04	B	1	91,8	0,42	cl1_14	B	0	9,8	-0,31
cl1_04	C	0	3,8	-0,26	cl1_14	C	1	69,7	0,47
cl1_04	M	0	0,2	-0,09	cl1_14	M	0	0,4	-0,07
cl1_04	N	0	1,1	-0,20	cl1_14	N	0	1,9	-0,25
cl1_05	A	1	83,7	0,48	cl1_15	A	1	58,5	0,49
cl1_05	B	0	7,7	-0,30	cl1_15	B	0	15,8	-0,33
cl1_05	C	0	6,9	-0,28	cl1_15	C	0	22,5	-0,18
cl1_05	M	0	0,3	-0,09	cl1_15	M	0	0,3	-0,07
cl1_05	N	0	1,4	-0,17	cl1_15	N	0	2,9	-0,25
cl1_06	A	0	11,4	-0,26	cl1_16	A	0	15,4	-0,15
cl1_06	B	0	7,2	-0,28	cl1_16	B	1	63,7	0,41
cl1_06	C	1	79,9	0,45	cl1_16	C	0	17,3	-0,26
cl1_06	M	0	0,4	-0,07	cl1_16	M	0	0,3	-0,07
cl1_06	N	0	1,1	-0,21	cl1_16	N	0	3,2	-0,24
cl1_07	A	1	86,2	0,47	cl1_17	A	1	68,6	0,50
cl1_07	B	0	7,1	-0,28	cl1_17	B	0	10,3	-0,31
cl1_07	C	0	5,0	-0,27	cl1_17	C	0	17,2	-0,23
cl1_07	M	0	0,3	-0,08	cl1_17	M	0	0,3	-0,07
cl1_07	N	0	1,4	-0,22	cl1_17	N	0	3,5	-0,24
cl1_08	A	0	7,1	-0,24	cl1_18	A	0	16,0	-0,22
cl1_08	B	0	14,1	-0,23	cl1_18	B	0	14,4	-0,28
cl1_08	C	1	76,2	0,41	cl1_18	C	1	65,6	0,48
cl1_08	M	0	0,3	-0,07	cl1_18	M	0	0,3	-0,07
cl1_08	N	0	2,3	-0,20	cl1_18	N	0	3,6	-0,25
cl1_09	A	1	80,1	0,43	cl1_19	A	0	16,7	-0,21
cl1_09	B	0	9,8	-0,30	cl1_19	B	1	58,3	0,45
cl1_09	C	0	8,4	-0,20	cl1_19	C	0	20,6	-0,22
cl1_09	M	0	0,5	-0,07	cl1_19	M	0	0,3	-0,07
cl1_09	N	0	1,3	-0,21	cl1_19	N	0	4,1	-0,24
cl1_10	A	0	17,6	-0,24	cl1_20	A	0	8,0	-0,22
cl1_10	B	0	25,6	-0,15	cl1_20	B	1	80,7	0,45
cl1_10	C	1	54,5	0,39	cl1_20	C	0	7,9	-0,25
cl1_10	M	0	0,4	-0,08	cl1_20	M	0	0,3	-0,07
cl1_10	N	0	1,9	-0,21	cl1_20	N	0	3,1	-0,26

Item	Alternativa	Valor	%	ptme
cl1_21	A	0	15,3	-0,12
cl1_21	B	1	45,9	0,39
cl1_21	C	0	34,1	-0,20
cl1_21	M	0	0,3	-0,07
cl1_21	N	0	4,4	-0,24
cl1_23	A	1	39,8	0,39
cl1_23	B	0	39,1	-0,11
cl1_23	C	0	16,3	-0,23
cl1_23	M	0	0,3	-0,07
cl1_23	N	0	4,6	-0,24
cl2_01	A	0	2,8	-0,17
cl2_01	B	1	91,0	0,33
cl2_01	C	0	5,1	-0,24
cl2_01	M	0	0,3	-0,08
cl2_01	N	0	0,7	-0,12
cl2_02	A	0	9,2	-0,25
cl2_02	B	0	6,2	-0,35
cl2_02	C	1	83,4	0,45
cl2_02	M	0	0,2	-0,06
cl2_02	N	0	0,9	-0,10
cl2_03	A	1	89,6	0,44
cl2_03	B	0	5,0	-0,32
cl2_03	C	0	4,8	-0,23
cl2_03	M	0	0,1	-0,06
cl2_03	N	0	0,5	-0,16
cl2_04	A	0	3,3	-0,23
cl2_04	B	1	91,8	0,41
cl2_04	C	0	4,1	-0,28
cl2_04	M	0	0,1	-0,06
cl2_04	N	0	0,6	-0,16
cl2_05	A	0	7,6	-0,25
cl2_05	B	1	84,2	0,44
cl2_05	C	0	7,2	-0,30
cl2_05	M	0	0,1	-0,07
cl2_05	N	0	0,9	-0,13
cl2_06	A	1	86,3	0,48
cl2_06	B	0	7,7	-0,33
cl2_06	C	0	5,0	-0,27
cl2_06	M	0	0,2	-0,06
cl2_06	N	0	0,9	-0,16
cl2_07	A	0	11,8	-0,26
cl2_07	B	0	10,5	-0,32
cl2_07	C	1	76,3	0,48
cl2_07	M	0	0,2	-0,06
cl2_07	N	0	1,2	-0,17
cl2_08	A	1	60,6	0,42
cl2_08	B	0	30,6	-0,26
cl2_08	C	0	7,7	-0,26
cl2_08	M	0	0,2	-0,06
cl2_08	N	0	0,9	-0,15
cl2_09	A	0	30,5	-0,22
cl2_09	B	1	56,9	0,40
cl2_09	C	0	11,1	-0,25
cl2_09	M	0	0,2	-0,06
cl2_09	N	0	1,3	-0,15
cl2_10	A	0	27,3	-0,20
cl2_10	B	0	23,8	-0,25
cl2_10	C	1	47,5	0,43
cl2_10	M	0	0,2	-0,05
cl2_10	N	0	1,2	-0,16
cl2_11	A	0	13,9	-0,25
cl2_11	B	1	69,5	0,43
cl2_11	C	0	15,0	-0,25
cl2_11	M	0	0,2	-0,06
cl2_11	N	0	1,4	-0,16

Item	Alternativa	Valor	%	ptme
cl2_12	A	0	13,1	-0,24
cl2_12	B	0	25,0	-0,24
cl2_12	C	1	59,7	0,42
cl2_12	M	0	0,3	-0,07
cl2_12	N	0	1,8	-0,14
cl2_13	A	0	18,7	-0,30
cl2_13	B	0	13,8	-0,34
cl2_13	C	1	66,1	0,54
cl2_13	M	0	0,3	-0,06
cl2_13	N	0	1,2	-0,17
cl2_14	A	1	72,9	0,47
cl2_14	B	0	9,4	-0,34
cl2_14	C	0	15,7	-0,23
cl2_14	M	0	0,2	-0,05
cl2_14	N	0	1,8	-0,17
cl2_15	A	1	79,0	0,49
cl2_15	B	0	11,4	-0,28
cl2_15	C	0	7,8	-0,31
cl2_15	M	0	0,2	-0,06
cl2_15	N	0	1,7	-0,19
cl2_16	A	0	23,0	-0,21
cl2_16	B	1	66,5	0,40
cl2_16	C	0	8,2	-0,26
cl2_16	M	0	0,2	-0,05
cl2_16	N	0	2,1	-0,17
cl2_17	A	0	9,5	-0,28
cl2_17	B	0	25,7	-0,30
cl2_17	C	1	62,4	0,49
cl2_17	M	0	0,2	-0,06
cl2_17	N	0	2,2	-0,17
cl2_18	A	0	11,6	-0,24
cl2_18	B	1	73,0	0,46
cl2_18	C	0	12,8	-0,29
cl2_18	M	0	0,2	-0,06
cl2_18	N	0	2,4	-0,17
cl2_19	A	0	14,4	-0,29
cl2_19	B	0	17,9	-0,27
cl2_19	C	1	65,6	0,49
cl2_19	M	0	0,2	-0,05
cl2_19	N	0	1,8	-0,18
cl2_20	A	1	71,7	0,52
cl2_20	B	0	14,6	-0,32
cl2_20	C	0	11,4	-0,30
cl2_20	M	0	0,2	-0,05
cl2_20	N	0	2,1	-0,19
cl2_21	A	0	14,2	-0,25
cl2_21	B	0	15,3	-0,21
cl2_21	C	1	67,8	0,42
cl2_21	M	0	0,2	-0,05
cl2_21	N	0	2,5	-0,18
cl2_22	A	1	62,4	0,45
cl2_22	B	0	16,0	-0,30
cl2_22	C	0	18,8	-0,19
cl2_22	M	0	0,2	-0,06
cl2_22	N	0	2,7	-0,18
cl2_23	A	0	16,8	-0,18
cl2_23	B	0	13,6	-0,30
cl2_23	C	1	66,3	0,42
cl2_23	M	0	0,2	-0,05
cl2_23	N	0	3,1	-0,17

## Anexo 10

## Porcentaje de selección y correlación alternativa - Medida de la prueba de comprensión lectora. MC-2013 segundo grado

Tabla 15 Características de las alternativas de la prueba de Comprensión Lectora MC-2013

Item	Alternativa	Valor	%	ptme	Item	Alternativa	Valor	%	ptme
cl1_01	A	0	2,5	-0,19	cl1_12	A	1	77,5	0,38
cl1_01	B	0	4,4	-0,27	cl1_12	B	0	16,4	-0,21
cl1_01	C	1	92,2	0,36	cl1_12	C	0	4,5	-0,26
cl1_01	M	0	0,3	-0,07	cl1_12	M	0	0,2	-0,06
cl1_01	N	0	0,5	-0,12	cl1_12	N	0	1,4	-0,22
cl1_02	A	0	8,2	-0,27	cl1_13	A	0	9,4	-0,21
cl1_02	B	1	87,9	0,40	cl1_13	B	1	68,0	0,37
cl1_02	C	0	3,1	-0,24	cl1_13	C	0	20,7	-0,21
cl1_02	M	0	0,3	-0,07	cl1_13	M	0	0,1	-0,06
cl1_02	N	0	0,6	-0,14	cl1_13	N	0	1,7	-0,21
cl1_03	A	1	87,3	0,45	cl1_14	A	0	18,5	-0,23
cl1_03	B	0	6,7	-0,33	cl1_14	B	0	10,1	-0,31
cl1_03	C	0	5,3	-0,23	cl1_14	C	1	69,6	0,47
cl1_03	M	0	0,2	-0,07	cl1_14	M	0	0,2	-0,06
cl1_03	N	0	0,6	-0,17	cl1_14	N	0	1,6	-0,23
cl1_04	A	0	3,0	-0,24	cl1_15	A	1	58,6	0,50
cl1_04	B	1	92,3	0,42	cl1_15	B	0	16,4	-0,34
cl1_04	C	0	3,8	-0,28	cl1_15	C	0	22,5	-0,19
cl1_04	M	0	0,1	-0,07	cl1_15	M	0	0,2	-0,06
cl1_04	N	0	0,8	-0,18	cl1_15	N	0	2,4	-0,23
cl1_05	A	1	84,0	0,48	cl1_16	A	0	15,5	-0,16
cl1_05	B	0	7,7	-0,30	cl1_16	B	1	62,8	0,42
cl1_05	C	0	7,1	-0,31	cl1_16	C	0	18,8	-0,27
cl1_05	M	0	0,1	-0,07	cl1_16	M	0	0,2	-0,06
cl1_05	N	0	1,2	-0,16	cl1_16	N	0	2,7	-0,22
cl1_06	A	0	11,7	-0,28	cl1_17	A	1	68,0	0,49
cl1_06	B	0	7,3	-0,29	cl1_17	B	0	10,7	-0,32
cl1_06	C	1	79,9	0,46	cl1_17	C	0	18,1	-0,23
cl1_06	M	0	0,2	-0,06	cl1_17	M	0	0,2	-0,06
cl1_06	N	0	0,8	-0,19	cl1_17	N	0	3,1	-0,23
cl1_07	A	1	86,7	0,48	cl1_18	A	0	16,9	-0,22
cl1_07	B	0	7,1	-0,30	cl1_18	B	0	15,6	-0,28
cl1_07	C	0	4,9	-0,29	cl1_18	C	1	64,1	0,48
cl1_07	M	0	0,1	-0,06	cl1_18	M	0	0,2	-0,06
cl1_07	N	0	1,1	-0,19	cl1_18	N	0	3,2	-0,23
cl1_08	A	0	7,1	-0,25	cl1_19	A	0	17,9	-0,20
cl1_08	B	0	14,4	-0,23	cl1_19	B	1	56,3	0,44
cl1_08	C	1	76,7	0,41	cl1_19	C	0	22,0	-0,22
cl1_08	M	0	0,1	-0,06	cl1_19	M	0	0,2	-0,06
cl1_08	N	0	1,7	-0,18	cl1_19	N	0	3,7	-0,23
cl1_09	A	1	79,9	0,43	cl1_20	A	0	8,6	-0,23
cl1_09	B	0	9,9	-0,31	cl1_20	B	1	79,9	0,45
cl1_09	C	0	8,9	-0,20	cl1_20	C	0	8,4	-0,26
cl1_09	M	0	0,2	-0,06	cl1_20	M	0	0,1	-0,06
cl1_09	N	0	1,1	-0,19	cl1_20	N	0	3,0	-0,25
cl1_10	A	0	18,5	-0,24	cl1_21	A	0	16,6	-0,11
cl1_10	B	0	27,2	-0,13	cl1_21	B	1	43,5	0,37
cl1_10	C	1	52,5	0,36	cl1_21	C	0	35,7	-0,20
cl1_10	M	0	0,2	-0,06	cl1_21	M	0	0,2	-0,05
cl1_10	N	0	1,6	-0,20	cl1_21	N	0	4,1	-0,23
cl1_11	A	1	70,4	0,36	cl1_23	A	1	38,8	0,38
cl1_11	B	0	17,3	-0,24	cl1_23	B	0	39,7	-0,09
cl1_11	C	0	10,9	-0,15	cl1_23	C	0	17,1	-0,24
cl1_11	M	0	0,1	-0,06	cl1_23	M	0	0,1	-0,06
cl1_11	N	0	1,2	-0,22	cl1_23	N	0	4,3	-0,23

Item	Alternativa	Valor	%	ptme
cl2_01	A	1	93,9	0,37
cl2_01	B	0	2,5	-0,22
cl2_01	C	0	3,1	-0,27
cl2_01	M	0	0,2	-0,08
cl2_01	N	0	0,3	-0,09
cl2_02	A	1	91,6	0,41
cl2_02	B	0	5,5	-0,32
cl2_02	C	0	2,5	-0,21
cl2_02	M	0	0,1	-0,05
cl2_02	N	0	0,4	-0,09
cl2_03	A	0	31,1	-0,16
cl2_03	B	1	55,4	0,33
cl2_03	C	0	12,5	-0,23
cl2_03	M	0	0,1	-0,05
cl2_03	N	0	0,9	-0,10
cl2_04	A	0	9,6	-0,28
cl2_04	B	1	86,7	0,39
cl2_04	C	0	2,8	-0,24
cl2_04	M	0	0,1	-0,04
cl2_04	N	0	0,8	-0,10
cl2_05	A	1	86,7	0,48
cl2_05	B	0	6,9	-0,34
cl2_05	C	0	5,3	-0,28
cl2_05	M	0	0,1	-0,05
cl2_05	N	0	1,0	-0,11
cl2_06	A	0	7,1	-0,28
cl2_06	B	1	67,9	0,41
cl2_06	C	0	23,6	-0,26
cl2_06	M	0	0,1	-0,05
cl2_06	N	0	1,3	-0,10
cl2_07	A	0	4,6	-0,26
cl2_07	B	0	4,6	-0,30
cl2_07	C	1	90,4	0,43
cl2_07	M	0	0,1	-0,05
cl2_07	N	0	0,3	-0,13
cl2_08	A	1	88,5	0,48
cl2_08	B	0	6,0	-0,33
cl2_08	C	0	4,8	-0,29
cl2_08	M	0	0,1	-0,05
cl2_08	N	0	0,5	-0,13
cl2_09	A	0	5,4	-0,28
cl2_09	B	1	86,4	0,46
cl2_09	C	0	7,3	-0,33
cl2_09	M	0	0,1	-0,05
cl2_09	N	0	0,8	-0,09
cl2_10	A	0	14,4	-0,25
cl2_10	B	1	67,1	0,46
cl2_10	C	0	17,8	-0,30
cl2_10	M	0	0,1	-0,04
cl2_10	N	0	0,6	-0,13
cl2_11	A	1	69,1	0,50
cl2_11	B	0	14,6	-0,33
cl2_11	C	0	15,1	-0,28
cl2_11	M	0	0,1	-0,04
cl2_11	N	0	1,0	-0,13
cl2_12	A	0	15,4	-0,24
cl2_12	B	0	10,8	-0,29
cl2_12	C	1	72,5	0,43
cl2_12	M	0	0,1	-0,04
cl2_12	N	0	1,1	-0,13
cl2_13	A	0	15,8	-0,25
cl2_13	B	0	14,0	-0,31
cl2_13	C	1	68,3	0,46
cl2_13	M	0	0,1	-0,05
cl2_13	N	0	1,8	-0,12

Item	Alternativa	Valor	%	ptme
cl2_15	A	1	70,8	0,51
cl2_15	B	0	10,7	-0,31
cl2_15	C	0	16,8	-0,32
cl2_15	M	0	0,1	-0,04
cl2_15	N	0	1,7	-0,12
cl2_16	A	0	11,1	-0,31
cl2_16	B	0	10,3	-0,34
cl2_16	C	1	77,5	0,52
cl2_16	M	0	0,1	-0,04
cl2_16	N	0	0,9	-0,15
cl2_17	A	1	48,4	0,38
cl2_17	B	0	21,0	-0,27
cl2_17	C	0	29,2	-0,14
cl2_17	M	0	0,1	-0,03
cl2_17	N	0	1,2	-0,14
cl2_18	A	0	21,6	-0,13
cl2_18	B	1	52,5	0,39
cl2_18	C	0	24,4	-0,29
cl2_18	M	0	0,1	-0,03
cl2_18	N	0	1,3	-0,14
cl2_19	A	0	30,3	-0,33
cl2_19	B	0	8,9	-0,32
cl2_19	C	1	59,4	0,53
cl2_19	M	0	0,1	-0,04
cl2_19	N	0	1,3	-0,15
cl2_20	A	0	10,9	-0,29
cl2_20	B	0	17,7	-0,35
cl2_20	C	1	69,8	0,54
cl2_20	M	0	0,1	-0,04
cl2_20	N	0	1,5	-0,16
cl2_21	A	1	50,8	0,44
cl2_21	B	0	22,3	-0,26
cl2_21	C	0	24,9	-0,20
cl2_21	M	0	0,1	-0,04
cl2_21	N	0	1,9	-0,15
cl2_22	A	0	14,8	-0,24
cl2_22	B	1	73,9	0,42
cl2_22	C	0	9,4	-0,26
cl2_22	M	0	0,1	-0,04
cl2_22	N	0	1,8	-0,16
cl2_23	A	1	54,9	0,37
cl2_23	B	0	20,9	-0,20
cl2_23	C	0	21,5	-0,20
cl2_23	M	0	0,1	-0,04
cl2_23	N	0	2,6	-0,12

## Anexo 11

## Porcentaje de selección y correlación alternativa - Medida de la prueba de matemática. ECE 2013 segundo grado

Tabla 16 Características de las alternativas de la prueba de Matemática ECE 2013

Item	Alternativa	Valor	%	ptme	Item	Alternativa	Valor	%	ptme
ma1_01	A	0	2,4	-0,14	ma1_14	A	0	5,8	-0,15
ma1_01	B	0	4,5	-0,25	ma1_14	B	1	78,0	0,40
ma1_01	C	1	91,9	0,33	ma1_14	C	0	13,5	-0,29
ma1_01	M	0	0,2	-0,06	ma1_14	M	0	0,2	-0,06
ma1_01	N	0	1,1	-0,14	ma1_14	N	0	2,4	-0,19
ma1_02	A	0	6,3	-0,19	ma1_15	A	1	72,1	0,55
ma1_02	B	1	84,6	0,38	ma1_15	B	0	5,7	-0,24
ma1_02	C	0	7,7	-0,27	ma1_15	C	0	18,7	-0,39
ma1_02	M	0	0,2	-0,06	ma1_15	M	0	0,3	-0,04
ma1_02	N	0	1,1	-0,13	ma1_15	N	0	3,2	-0,20
ma1_04	A	1	72,0	0,46	ma1_16	A	0	67,5	-0,41
ma1_04	B	0	18,2	-0,30	ma1_16	B	0	3,0	-0,14
ma1_04	C	0	7,0	-0,22	ma1_16	C	1	26,9	0,55
ma1_04	M	0	0,4	-0,07	ma1_16	M	0	0,4	-0,03
ma1_04	N	0	2,4	-0,18	ma1_16	N	0	2,1	-0,18
ma1_05	A	0	8,5	-0,23	ma1_17	A	1	57,6	0,57
ma1_05	B	0	9,9	-0,28	ma1_17	B	0	13,9	-0,21
ma1_05	C	1	79,4	0,42	ma1_17	C	0	24,7	-0,40
ma1_05	M	0	0,5	-0,07	ma1_17	M	0	0,3	-0,05
ma1_05	N	0	1,6	-0,15	ma1_17	N	0	3,5	-0,20
ma1_06	A	1	78,0	0,45	ma1_18	A	0	31,1	-0,36
ma1_06	B	0	15,2	-0,35	ma1_18	B	1	52,1	0,54
ma1_06	C	0	4,8	-0,18	ma1_18	C	0	11,3	-0,21
ma1_06	M	0	0,6	-0,08	ma1_18	M	0	0,3	-0,05
ma1_06	N	0	1,4	-0,15	ma1_18	N	0	5,2	-0,17
ma1_07	A	0	17,0	-0,29	ma1_19	A	0	22,5	-0,38
ma1_07	B	1	58,2	0,45	ma1_19	B	1	52,8	0,49
ma1_07	C	0	21,4	-0,20	ma1_19	C	0	20,5	-0,11
ma1_07	M	0	0,3	-0,06	ma1_19	M	0	0,3	-0,05
ma1_07	N	0	3,1	-0,15	ma1_19	N	0	3,9	-0,20
ma1_08	A	0	12,3	-0,26	ma1_20	A	0	24,0	-0,40
ma1_08	B	0	23,8	-0,40	ma1_20	B	0	7,3	-0,25
ma1_08	C	1	62,0	0,57	ma1_20	C	1	64,2	0,59
ma1_08	M	0	0,4	-0,05	ma1_20	M	0	0,3	-0,05
ma1_08	N	0	1,6	-0,17	ma1_20	N	0	4,3	-0,20
ma1_09	A	0	12,3	-0,31	ma1_21	A	1	48,3	0,55
ma1_09	B	0	11,9	-0,26	ma1_21	B	0	16,1	-0,18
ma1_09	C	1	73,8	0,48	ma1_21	C	0	29,3	-0,35
ma1_09	M	0	0,3	-0,06	ma1_21	M	0	0,3	-0,06
ma1_09	N	0	1,7	-0,17	ma1_21	N	0	6,0	-0,20
ma1_11	A	1	67,7	0,58	ma2_01	A	0	6,7	-0,27
ma1_11	B	0	16,9	-0,38	ma2_01	B	0	7,2	-0,23
ma1_11	C	0	13,1	-0,29	ma2_01	C	1	85,4	0,39
ma1_11	M	0	0,2	-0,05	ma2_01	M	0	0,2	-0,05
ma1_11	N	0	2,0	-0,19	ma2_01	N	0	0,6	-0,11
ma1_12	A	0	40,6	-0,35	ma2_03	A	1	67,3	0,41
ma1_12	B	0	17,2	-0,21	ma2_03	B	0	21,5	-0,27
ma1_12	C	1	39,4	0,58	ma2_03	C	0	9,7	-0,21
ma1_12	M	0	0,3	-0,05	ma2_03	M	0	0,2	-0,05
ma1_12	N	0	2,5	-0,17	ma2_03	N	0	1,3	-0,14
ma1_13	A	0	25,5	-0,33	ma2_04	A	0	12,8	-0,33
ma1_13	B	1	58,1	0,51	ma2_04	B	1	78,0	0,43
ma1_13	C	0	13,9	-0,23	ma2_04	C	0	7,2	-0,19
ma1_13	M	0	0,2	-0,05	ma2_04	M	0	0,2	-0,05
ma1_13	N	0	2,3	-0,19	ma2_04	N	0	1,8	-0,15

Item	Alternativa	Valor	%	ptme
ma2_05	A	0	5,0	-0,24
ma2_05	B	0	7,7	-0,25
ma2_05	C	1	85,6	0,39
ma2_05	M	0	0,1	-0,05
ma2_05	N	0	1,7	-0,15
ma2_06	A	1	87,6	0,39
ma2_06	B	0	4,1	-0,23
ma2_06	C	0	7,1	-0,25
ma2_06	M	0	0,2	-0,06
ma2_06	N	0	1,0	-0,15
ma2_07	A	0	9,9	-0,29
ma2_07	B	0	14,1	-0,37
ma2_07	C	1	74,9	0,54
ma2_07	M	0	0,2	-0,05
ma2_07	N	0	0,9	-0,15
ma2_08	A	0	1,3	-0,07
ma2_08	B	1	39,1	0,60
ma2_08	C	0	58,8	-0,55
ma2_08	M	0	0,2	-0,03
ma2_08	N	0	0,7	-0,14
ma2_09	A	0	19,3	-0,37
ma2_09	B	0	26,7	-0,27
ma2_09	C	1	52,0	0,57
ma2_09	M	0	0,2	-0,04
ma2_09	N	0	1,8	-0,12
ma2_10	A	1	75,1	0,54
ma2_10	B	0	10,0	-0,33
ma2_10	C	0	13,8	-0,35
ma2_10	M	0	0,2	-0,05
ma2_10	N	0	0,8	-0,14
ma2_11	A	0	7,0	-0,23
ma2_11	B	1	85,5	0,39
ma2_11	C	0	6,2	-0,25
ma2_11	M	0	0,1	-0,05
ma2_11	N	0	1,1	-0,17
ma2_12	A	0	15,9	-0,35
ma2_12	B	1	69,8	0,52
ma2_12	C	0	10,9	-0,26
ma2_12	M	0	0,2	-0,04
ma2_12	N	0	3,3	-0,14
ma2_13	A	0	6,5	-0,14
ma2_13	B	0	39,6	-0,50
ma2_13	C	1	52,5	0,60
ma2_13	M	0	0,2	-0,04
ma2_13	N	0	1,2	-0,16
ma2_14	A	0	18,1	-0,26
ma2_14	B	1	66,9	0,50
ma2_14	C	0	12,9	-0,32
ma2_14	M	0	0,2	-0,05
ma2_14	N	0	1,9	-0,16
ma2_15	A	1	37,0	0,51
ma2_15	B	0	13,8	-0,05
ma2_15	C	0	47,3	-0,42
ma2_15	M	0	0,2	-0,03
ma2_15	N	0	1,7	-0,17
ma2_16	A	1	67,0	0,48
ma2_16	B	0	14,0	-0,21
ma2_16	C	0	15,6	-0,34
ma2_16	M	0	0,2	-0,04
ma2_16	N	0	3,3	-0,16
ma2_17	A	0	13,9	-0,14
ma2_17	B	1	55,6	0,55
ma2_17	C	0	27,9	-0,43
ma2_17	M	0	0,2	-0,04
ma2_17	N	0	2,5	-0,18

Item	Alternativa	Valor	%	ptme
ma2_18	A	1	55,0	0,59
ma2_18	B	0	13,0	-0,29
ma2_18	C	0	28,1	-0,37
ma2_18	M	0	0,2	-0,04
ma2_18	N	0	3,7	-0,14
ma2_19	A	0	25,9	-0,35
ma2_19	B	0	26,4	-0,31
ma2_19	C	1	45,3	0,63
ma2_19	M	0	0,2	-0,04
ma2_19	N	0	2,0	-0,18
ma2_20	A	1	43,3	0,43
ma2_20	B	0	18,4	-0,15
ma2_20	C	0	34,8	-0,26
ma2_20	M	0	0,2	-0,04
ma2_20	N	0	3,3	-0,17
ma2_21	A	0	24,8	-0,11
ma2_21	B	0	21,3	-0,34
ma2_21	C	1	50,7	0,43
ma2_21	M	0	0,2	-0,04
ma2_21	N	0	2,9	-0,18

## Anexo 12

## Porcentaje de selección y correlación alternativa - Medida de la prueba de matemática. MC-2013 segundo grado

Tabla 17 Características de las alternativas de la prueba de Matemática MC-2013

Item	Alternativa	Valor	%	ptme
ma1_01	A	0	2,4	-0,15
ma1_01	B	0	4,5	-0,25
ma1_01	C	1	92,3	0,32
ma1_01	M	0	0,1	-0,05
ma1_01	N	0	0,7	-0,11
ma1_02	A	0	6,7	-0,21
ma1_02	B	1	84,4	0,38
ma1_02	C	0	8,0	-0,28
ma1_02	M	0	0,1	-0,05
ma1_02	N	0	0,8	-0,11
ma1_04	A	1	71,6	0,46
ma1_04	B	0	19,3	-0,32
ma1_04	C	0	7,2	-0,23
ma1_04	M	0	0,2	-0,05
ma1_04	N	0	1,7	-0,14
ma1_05	A	0	9,2	-0,24
ma1_05	B	0	10,8	-0,27
ma1_05	C	1	78,3	0,42
ma1_05	M	0	0,3	-0,06
ma1_05	N	0	1,4	-0,12
ma1_06	A	1	77,3	0,45
ma1_06	B	0	16,1	-0,37
ma1_06	C	0	5,1	-0,17
ma1_06	M	0	0,3	-0,06
ma1_06	N	0	1,2	-0,11
ma1_07	A	0	18,5	-0,29
ma1_07	B	1	55,8	0,45
ma1_07	C	0	22,8	-0,21
ma1_07	M	0	0,1	-0,04
ma1_07	N	0	2,7	-0,12
ma1_08	A	0	13,8	-0,27
ma1_08	B	0	24,2	-0,40
ma1_08	C	1	60,6	0,57
ma1_08	M	0	0,2	-0,04
ma1_08	N	0	1,3	-0,14
ma1_09	A	0	12,6	-0,32
ma1_09	B	0	12,3	-0,26
ma1_09	C	1	73,6	0,48
ma1_09	M	0	0,1	-0,04
ma1_09	N	0	1,4	-0,13
ma1_11	A	1	65,9	0,57
ma1_11	B	0	17,5	-0,39
ma1_11	C	0	14,5	-0,28
ma1_11	M	0	0,1	-0,04
ma1_11	N	0	1,9	-0,16
ma1_12	A	0	43,4	-0,35
ma1_12	B	0	16,9	-0,22
ma1_12	C	1	37,0	0,58
ma1_12	M	0	0,1	-0,04
ma1_12	N	0	2,5	-0,15
ma1_13	A	0	27,2	-0,33
ma1_13	B	1	56,4	0,51
ma1_13	C	0	14,1	-0,24
ma1_13	M	0	0,1	-0,04
ma1_13	N	0	2,2	-0,15
ma1_14	A	0	5,6	-0,14
ma1_14	B	1	77,9	0,39
ma1_14	C	0	14,0	-0,30
ma1_14	M	0	0,1	-0,04
ma1_14	N	0	2,3	-0,16
ma1_15	A	1	71,6	0,55
ma1_15	B	0	5,3	-0,23
ma1_15	C	0	20,0	-0,41
ma1_15	M	0	0,1	-0,04
ma1_15	N	0	2,9	-0,18
ma1_16	A	0	70,6	-0,44
ma1_16	B	0	2,7	-0,13
ma1_16	C	1	24,6	0,57
ma1_16	M	0	0,2	-0,03
ma1_16	N	0	2,0	-0,17
ma1_17	A	1	58,5	0,58
ma1_17	B	0	13,2	-0,21
ma1_17	C	0	25,1	-0,42
ma1_17	M	0	0,1	-0,04
ma1_17	N	0	3,1	-0,18
ma1_18	A	0	33,2	-0,36
ma1_18	B	1	50,1	0,55
ma1_18	C	0	11,6	-0,22
ma1_18	M	0	0,1	-0,04
ma1_18	N	0	5,0	-0,15
ma1_19	A	0	22,9	-0,39
ma1_19	B	1	51,9	0,49
ma1_19	C	0	21,2	-0,11
ma1_19	M	0	0,2	-0,04
ma1_19	N	0	3,9	-0,18
ma1_20	A	0	26,1	-0,41
ma1_20	B	0	7,0	-0,25
ma1_20	C	1	62,5	0,58
ma1_20	M	0	0,1	-0,04
ma1_20	N	0	4,2	-0,18
ma1_21	A	1	47,3	0,55
ma1_21	B	0	16,4	-0,17
ma1_21	C	0	30,3	-0,36
ma1_21	M	0	0,2	-0,05
ma1_21	N	0	5,9	-0,19
ma2_01	A	0	4,5	-0,22
ma2_01	B	0	15,4	-0,27
ma2_01	C	1	79,4	0,38
ma2_01	M	0	0,1	-0,04
ma2_01	N	0	0,6	-0,10
ma2_03	A	0	6,9	-0,21
ma2_03	B	1	84,0	0,39
ma2_03	C	0	8,4	-0,30
ma2_03	M	0	0,1	-0,04
ma2_03	N	0	0,6	-0,10
ma2_05	A	0	5,5	-0,20
ma2_05	B	0	6,5	-0,25
ma2_05	C	1	86,8	0,36
ma2_05	M	0	0,1	-0,04
ma2_05	N	0	1,1	-0,13

Item	Alternativa	Valor	%	ptme
ma2_06	A	0	5,7	-0,24
ma2_06	B	0	7,7	-0,27
ma2_06	C	1	85,0	0,41
ma2_06	M	0	0,1	-0,04
ma2_06	N	0	1,5	-0,13
ma2_07	A	1	62,4	0,56
ma2_07	B	0	12,4	-0,14
ma2_07	C	0	24,2	-0,49
ma2_07	M	0	0,1	-0,04
ma2_07	N	0	0,9	-0,13
ma2_08	A	0	16,5	-0,40
ma2_08	B	0	9,8	-0,27
ma2_08	C	1	71,9	0,55
ma2_08	M	0	0,1	-0,03
ma2_08	N	0	1,7	-0,14
ma2_09	A	0	17,1	-0,42
ma2_09	B	1	51,9	0,53
ma2_09	C	0	29,4	-0,19
ma2_09	M	0	0,1	-0,03
ma2_09	N	0	1,5	-0,15
ma2_10	A	0	8,3	-0,29
ma2_10	B	1	67,2	0,49
ma2_10	C	0	23,6	-0,33
ma2_10	M	0	0,1	-0,04
ma2_10	N	0	0,8	-0,13
ma2_11	A	0	62,1	-0,56
ma2_11	B	0	2,3	-0,11
ma2_11	C	1	34,7	0,63
ma2_11	M	0	0,1	-0,02
ma2_11	N	0	0,7	-0,13
ma2_12	A	1	55,9	0,59
ma2_12	B	0	8,8	-0,28
ma2_12	C	0	32,5	-0,41
ma2_12	M	0	0,1	-0,03
ma2_12	N	0	2,7	-0,13
ma2_13	A	1	54,3	0,59
ma2_13	B	0	11,7	-0,08
ma2_13	C	0	32,8	-0,53
ma2_13	M	0	0,1	-0,03
ma2_13	N	0	1,1	-0,14
ma2_14	A	0	34,1	-0,43
ma2_14	B	1	36,2	0,51
ma2_14	C	0	27,7	-0,05
ma2_14	M	0	0,1	-0,03
ma2_14	N	0	1,9	-0,13
ma2_15	A	0	32,2	-0,47
ma2_15	B	0	12,9	-0,21
ma2_15	C	1	53,3	0,62
ma2_15	M	0	0,1	-0,03
ma2_15	N	0	1,5	-0,15
ma2_16	A	1	75,9	0,48
ma2_16	B	0	12,5	-0,28
ma2_16	C	0	9,8	-0,31
ma2_16	M	0	0,1	-0,03
ma2_16	N	0	1,7	-0,16
ma2_17	A	0	24,1	-0,38
ma2_17	B	1	49,9	0,52
ma2_17	C	0	23,2	-0,17
ma2_17	M	0	0,1	-0,03
ma2_17	N	0	2,7	-0,15
ma2_18	A	0	6,5	-0,10
ma2_18	B	1	64,9	0,59
ma2_18	C	0	27,0	-0,53
ma2_18	M	0	0,1	-0,02
ma2_18	N	0	1,6	-0,16

Item	Alternativa	Valor	%	ptme
ma2_19	A	0	31,0	-0,21
ma2_19	B	0	16,4	-0,28
ma2_19	C	1	50,2	0,45
ma2_19	M	0	0,1	-0,03
ma2_19	N	0	2,2	-0,15
ma2_20	A	1	42,8	0,62
ma2_20	B	0	15,4	-0,26
ma2_20	C	0	39,3	-0,38
ma2_20	M	0	0,1	-0,02
ma2_20	N	0	2,5	-0,17
ma2_21	A	0	32,6	-0,32
ma2_21	B	1	51,4	0,44
ma2_21	C	0	13,5	-0,13
ma2_21	M	0	0,1	-0,03
ma2_21	N	0	2,4	-0,16



## Anexo 13

## Porcentaje de selección y correlación alternativa - Medida de la prueba de comprensión lectora - castellano como segunda lengua. ECE 2013 cuarto grado EIB

**Tabla 18** Características de las alternativas de la prueba de Comprensión Lectora – Castellano como segunda lengua ECE 2013 cuarto grado EIB.

Item	Alternativa	Valor	%	ptme	Item	Alternativa	Valor	%	ptme
L2_1_01	A	1	78,0	0,45	L2_1_12	A	0	14,7	-0,11
L2_1_01	B	0	10,0	-0,28	L2_1_12	B	1	59,5	0,30
L2_1_01	C	0	7,4	-0,20	L2_1_12	C	0	21,8	-0,12
L2_1_01	M	0	0,7	-0,12	L2_1_12	M	0	0,7	-0,12
L2_1_01	N	0	3,9	-0,22	L2_1_12	N	0	3,3	-0,28
L2_1_02	A	0	4,7	-0,20	L2_1_13	A	0	15,8	-0,10
L2_1_02	B	0	10,9	-0,26	L2_1_13	B	1	46,6	0,46
L2_1_02	C	1	82,7	0,41	L2_1_13	C	0	32,7	-0,27
L2_1_02	M	0	0,7	-0,16	L2_1_13	M	0	0,6	-0,12
L2_1_02	N	0	1,0	-0,18	L2_1_13	N	0	4,3	-0,28
L2_1_03	A	0	5,1	-0,18	L2_1_14	A	0	24,0	-0,15
L2_1_03	B	1	85,4	0,41	L2_1_14	B	0	17,9	-0,23
L2_1_03	C	0	6,9	-0,25	L2_1_14	C	1	52,9	0,44
L2_1_03	M	0	0,5	-0,14	L2_1_14	M	0	0,7	-0,13
L2_1_03	N	0	2,0	-0,22	L2_1_14	N	0	4,5	-0,28
L2_1_04	A	0	15,9	-0,18	L2_1_15	A	1	40,6	0,41
L2_1_04	B	0	9,0	-0,29	L2_1_15	B	0	33,8	-0,22
L2_1_04	C	1	72,7	0,42	L2_1_15	C	0	20,4	-0,08
L2_1_04	M	0	0,8	-0,14	L2_1_15	M	0	0,5	-0,13
L2_1_04	N	0	1,6	-0,22	L2_1_15	N	0	4,7	-0,28
L2_1_05	A	1	70,6	0,55	L2_1_16	A	0	22,7	-0,21
L2_1_05	B	0	16,2	-0,33	L2_1_16	B	1	61,8	0,47
L2_1_05	C	0	10,5	-0,28	L2_1_16	C	0	10,0	-0,22
L2_1_05	M	0	0,6	-0,13	L2_1_16	M	0	0,6	-0,12
L2_1_05	N	0	2,2	-0,24	L2_1_16	N	0	5,0	-0,29
L2_1_06	A	0	5,9	-0,20	L2_1_17	A	0	18,9	-0,15
L2_1_06	B	0	11,7	-0,25	L2_1_17	B	1	39,4	0,30
L2_1_06	C	1	80,1	0,43	L2_1_17	C	0	34,9	-0,02
L2_1_06	M	0	0,7	-0,14	L2_1_17	M	0	0,6	-0,12
L2_1_06	N	0	1,6	-0,26	L2_1_17	N	0	6,2	-0,28
L2_1_07	A	1	53,2	0,40	L2_1_18	A	1	62,5	0,56
L2_1_07	B	0	31,0	-0,20	L2_1_18	B	0	13,7	-0,26
L2_1_07	C	0	13,1	-0,19	L2_1_18	C	0	16,7	-0,26
L2_1_07	M	0	0,7	-0,13	L2_1_18	M	0	0,6	-0,13
L2_1_07	N	0	2,0	-0,24	L2_1_18	N	0	6,5	-0,29
L2_1_08	A	0	36,3	-0,06	L2_1_19	A	1	46,0	0,52
L2_1_08	B	1	47,1	0,30	L2_1_19	B	0	20,9	-0,25
L2_1_08	C	0	13,7	-0,21	L2_1_19	C	0	25,9	-0,18
L2_1_08	M	0	0,8	-0,14	L2_1_19	M	0	0,5	-0,12
L2_1_08	N	0	2,2	-0,25	L2_1_19	N	0	6,8	-0,28
L2_1_09	A	0	13,2	-0,23	L2_1_21	A	1	55,9	0,56
L2_1_09	B	0	13,4	-0,31	L2_1_21	B	0	20,1	-0,28
L2_1_09	C	1	70,4	0,51	L2_1_21	C	0	13,7	-0,19
L2_1_09	M	0	0,6	-0,14	L2_1_21	M	0	0,5	-0,12
L2_1_09	N	0	2,4	-0,26	L2_1_21	N	0	9,8	-0,30
L2_1_10	A	1	33,7	0,39	L2_1_22	A	1	55,1	0,52
L2_1_10	B	0	32,0	-0,25	L2_1_22	B	0	16,6	-0,23
L2_1_10	C	0	31,1	-0,04	L2_1_22	C	0	19,0	-0,22
L2_1_10	M	0	0,7	-0,11	L2_1_22	M	0	0,8	-0,13
L2_1_10	N	0	2,6	-0,27	L2_1_22	N	0	8,5	-0,26
L2_1_11	A	0	30,4	-0,07	L2_2_01	A	0	2,7	-0,12
L2_1_11	B	0	24,6	-0,17	L2_2_01	B	1	89,1	0,36
L2_1_11	C	1	41,5	0,32	L2_2_01	C	0	5,4	-0,25
L2_1_11	M	0	0,6	-0,13	L2_2_01	M	0	0,5	-0,11
L2_1_11	N	0	2,8	-0,26	L2_2_01	N	0	2,3	-0,19

Item	Alternativa	Valor	%	ptme
L2_2_02	A	1	81,4	0,48
L2_2_02	B	0	11,9	-0,34
L2_2_02	C	0	5,2	-0,22
L2_2_02	M	0	0,5	-0,16
L2_2_02	N	0	1,1	-0,17
L2_2_03	A	1	75,5	0,53
L2_2_03	B	0	13,1	-0,34
L2_2_03	C	0	9,4	-0,28
L2_2_03	M	0	0,6	-0,16
L2_2_03	N	0	1,3	-0,15
L2_2_04	A	1	54,1	0,40
L2_2_04	B	0	23,3	-0,19
L2_2_04	C	0	20,7	-0,21
L2_2_04	M	0	0,8	-0,13
L2_2_04	N	0	1,0	-0,19
L2_2_05	A	0	9,0	-0,21
L2_2_05	B	0	22,9	-0,35
L2_2_05	C	1	65,7	0,51
L2_2_05	M	0	0,6	-0,14
L2_2_05	N	0	1,8	-0,19
L2_2_06	A	1	65,3	0,47
L2_2_06	B	0	19,2	-0,24
L2_2_06	C	0	13,6	-0,28
L2_2_06	M	0	0,6	-0,15
L2_2_06	N	0	1,3	-0,20
L2_2_07	A	1	71,2	0,42
L2_2_07	B	0	9,5	-0,27
L2_2_07	C	0	17,1	-0,19
L2_2_07	M	0	0,7	-0,15
L2_2_07	N	0	1,4	-0,21
L2_2_10	A	1	25,3	0,29
L2_2_10	B	0	34,3	-0,21
L2_2_10	C	0	37,4	0,02
L2_2_10	M	0	0,6	-0,13
L2_2_10	N	0	2,4	-0,18
L2_2_11	A	1	52,4	0,46
L2_2_11	B	0	23,8	-0,23
L2_2_11	C	0	21,0	-0,22
L2_2_11	M	0	0,7	-0,13
L2_2_11	N	0	2,1	-0,21
L2_2_12	A	0	23,6	-0,20
L2_2_12	B	1	55,4	0,48
L2_2_12	C	0	18,5	-0,29
L2_2_12	M	0	0,6	-0,14
L2_2_12	N	0	2,0	-0,22
L2_2_13	A	0	16,7	-0,20
L2_2_13	B	0	26,1	-0,22
L2_2_13	C	1	54,1	0,43
L2_2_13	M	0	0,6	-0,13
L2_2_13	N	0	2,5	-0,21
L2_2_14	A	1	39,9	0,45
L2_2_14	B	0	29,1	-0,22
L2_2_14	C	0	27,8	-0,17
L2_2_14	M	0	0,6	-0,14
L2_2_14	N	0	2,6	-0,22
L2_2_15	A	1	45,1	0,51
L2_2_15	B	0	31,2	-0,26
L2_2_15	C	0	20,2	-0,23
L2_2_15	M	0	0,6	-0,14
L2_2_15	N	0	2,9	-0,21
L2_2_16	A	1	54,4	0,50
L2_2_16	B	0	28,9	-0,30
L2_2_16	C	0	13,4	-0,20
L2_2_16	M	0	0,5	-0,13
L2_2_16	N	0	2,8	-0,22

Item	Alternativa	Valor	%	ptme
L2_2_17	A	0	26,7	-0,07
L2_2_17	B	0	32,3	-0,19
L2_2_17	C	1	37,2	0,35
L2_2_17	M	0	0,6	-0,12
L2_2_17	N	0	3,2	-0,23
L2_2_18	A	0	26,0	-0,07
L2_2_18	B	1	36,8	0,34
L2_2_18	C	0	33,1	-0,18
L2_2_18	M	0	0,5	-0,13
L2_2_18	N	0	3,5	-0,22
L2_2_19	A	1	49,8	0,44
L2_2_19	B	0	24,9	-0,22
L2_2_19	C	0	21,2	-0,18
L2_2_19	M	0	0,5	-0,12
L2_2_19	N	0	3,6	-0,22
L2_2_20	A	0	22,5	-0,17
L2_2_20	B	0	20,1	-0,24
L2_2_20	C	1	53,1	0,43
L2_2_20	M	0	0,6	-0,11
L2_2_20	N	0	3,7	-0,21
L2_2_21	A	1	46,5	0,46
L2_2_21	B	0	26,8	-0,26
L2_2_21	C	0	22,0	-0,15
L2_2_21	M	0	0,5	-0,11
L2_2_21	N	0	4,1	-0,21



Ministerio de Educación del Perú  
Calle Del Comercio 193, San Borja  
Lima, Perú  
Teléfonos: 615-5800 / 615-5840

[www.minedu.gob.pe](http://www.minedu.gob.pe)